Judgement, Capabilities and Practice

Exploring how clinical supervisors make trust judgements and whether they can be captured by a capability model

Judgement, Capabilities and Practice

Exploring how clinical supervisors make trust judgements and whether they can be captured by a capability model

Doctor in Education (EdD)
Institute of Education, University College London

Maria Bussey

December 2018

Declaration

I confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Maria Bussey

1 March 2019

Word count of chapters 1-11 (including tables): 44,906

Abstract

The study had two aims; the first was to gain an insight into the patterns of judgement of clinical supervisors. The second was to explore how these patterns were altered by or corresponded with the use of a proposed assessment framework (Capabilities in Practice or CiPs) in order to help align assessment theory with practice.

Designed as a qualitative study, it involved two separate interviews with six consultant surgeons, from different backgrounds, specialties and regions. The first interview explored the strategies they used to make judgements about trainees. The second interview followed a simulation of CiPs and explored the differences in the way trainees were judged.

From thematic analysis, the findings from the first set of interviews showed that there were a number of important influences on supervisor judgement arising from political, cultural and financial factors in the organisational context (NHS practice). Drawing on the theories of judgement and trust, a new theory, Judgement in Action, was developed, showing that supervisor judgement had a particular pattern when applied to the judgement of trainee performance and progression. Judgement in Action incorporated four types of training capability under which there were qualities trainees were expected to exhibit when undertaking any high-level training activity. Two of these capabilities were core to surgery and two complementary to it. Two capabilities tended to be learned through instruction while the others were mainly learned through experience. The latter were higher-order, involving complex judgement and skill.

The study proposed using the theory of Judgement in Action to help improve the structure of curricula and assessment and to help supervisors and trainees reflect on judgement practices to improve performance. It also proposed that the theory of Judgement in Action could be explored in training regimes used by other professional groups.

Word count: 293

Impact statement

The core of this study was an analysis and demystification of the way consultant surgeons, acting as clinical supervisors, tended to make judgements about the capability of trainees. These judgements were crucial to the safety of patients undergoing surgical procedures in UK hospitals and were part of a process referred to in this study as Judgement in Action.

The theory of Judgement in Action described an active process, involving clinical supervisors' concrete experiences with trainees in different training situations. The conscious and unconscious knowledge about the trainee gained from these experiences fell into four categories; competence, decision-making, administration and relationship skills. Supervisor Judgements in Action were used to make decisions about the degree of supervision that was required, or conversely, the amount of trust that could be granted to trainees carrying out clinical work.

The adoption of the theory of Judgement in Action has the potential to improve surgical training, assessment and practice in relation to curriculum, supervisor judgement, learner understanding, patient safety and organisational culture. This would be at a local, regional, national or international level.

Curriculum: Judgement in Action, as part of a major curriculum revision, is being used to improve the design of assessments to ensure that they work more closely with the natural judgement patterns of consultant clinical supervisors. It has

helped to validate the importance of professional judgement in building a more holistic picture of trainee ability.

Supervisor judgement: Consultant clinical supervisors are likely to benefit from a better understanding of their natural judgement behaviour. Judgement in Action provides a framework upon which supervisors can reflect in order to make better judgements. In addition, if the theoretical model of assessment was more closely aligned with supervisor judgement-making, supervisors would be more likely to be convinced of its legitimacy and undertake it more conscientiously.

Learner understanding: If trainees had a better understanding of how they were judged, they would be better placed to ensure they achieved what was required of them.

Patient safety: The ultimate aim of surgical training is to provide patient-safe surgeons. Trainees need to be trained and assessed thoroughly, and ensuring trainees were proficient in all four areas of the Judgement in Action framework would help to provide a more holistic assessment with fewer gaps at the end of the training programme.

Organisational culture: Judgement in Action would affect how trainees were viewed in different training environments such as the operating theatre, on the ward and out-patient clinic. In each environment there would be equal prominence given to competence, sound decision-making, relationships with patients and colleagues and good organisational abilities.

Other medical and non-medical fields and professions: Judgement in Action might serve well as a practical approach to the judgement of learner performance in a variety of professions. While the four overarching capabilities would remain unchanged, the underlying features of those capabilities could be unique to a field of practice. For example, in airline industry training, recruits require technical expertise and knowledge of protocols (competence), the ability to communicate with and co-ordinate ground and flight teams (relationships), the ability to follow a documented process (administration) and make rapid decisions to cope with system and human errors as well as security threats (decision-making).

Word count: 531

Reflective statement

Introduction

I had a clear goal in mind as I first approached the Institute of Education on Saturday 4th October 2014 but questioned my ability to achieve it. I wanted to validate the employment of workplace-based assessment (WBA) in surgical training which appeared to be shrouded by misunderstanding and miscommunication. I was sure of three things; the first was that there was a need for qualitative research on WBA in surgical training, the second was that I aspired to being part of it and the third that, in working as a curriculum developer for the Royal Colleges of Surgeons, I was in a unique position from which to participate. WBA needed further development, however, qualitative real world research had not been employed in this area, and I had little idea how to initiate it.

Within a few weeks of the course I felt glad that I had chosen the EdD because of its structured nature and collaborative group support. The course provided me with reassurance that this area of focus was feasible and appropriate and potentially extremely valuable for my professional development. I felt the course was giving me the tools, while the inspiration came from observing the professional behaviour of the surgeons with whom I worked.

Reflecting on my career to date, I could see that I had defined myself as a manager with an MBA. While I had considerable experience in team-building, staff development and project management, I had reached an impasse that was

stultifying. I was looking for a challenge, more opportunities and to be able to add a new dimension to my career pathway.

Progression through taught courses

FOUNDATIONS OF PROFESSIONALISM (FOP)

I found the FOP the most challenging component of the EdD. It introduced me to a brand new lexicon of terminology. At first, I found this language and the manner of writing alien and asked myself whether these obfuscated or clarified. I soon realised that the terminology was a key that helped unlock different perspectives. I needed to engage with a broad literature by reading widely.

Through reading, I learned that qualitative and quantitative research bore little resemblance to each other and what it meant to take an interpretive, constructivist view that was part of the research itself. I learned that reflexivity was a crucial element of any qualitative research project, entailing awareness of how the researcher affects the subject matter being investigated and that any knowledge gained was therefore subject to their unique perspective. I learned how to undertake a literature review and how to constructively critique published work by asking *why*?, *what*? and *so what*? The readers, particularly Robson (2011) were very helpful to me in coming to terms with different concepts and relating them to my area of work. This was aided by approachable and supportive tutors and friendly fellow students, willing to share ideas and concepts.

The assignment gave me the opportunity to analyse the nature of professionalism in relation to the practice of surgery. I learned the importance of being able to step outside my professional role and to properly reflect on influences that shaped

surgery, thereby gaining an insight into the ways in which the performativity of curriculum co-existed with the nature of surgical practice. I felt that Schön's (1990, p.3) metaphor of the hard and swampy ground was apposite.

The feedback I received on the initial submission was very detailed and helpful. I learned that my established style of descriptive writing was inappropriate and after an initial crisis of confidence, realised the importance of using an analytical and conceptual framework. The tutor helpfully pointed me towards Freidson (1994) which gave me a lens for exposing how professionalism for surgeons was changing in the context of the new curriculum. I spent many hours over the Christmas break that would have normally been spent celebrating, huddled over my PC preparing the initial submission and again in the New Year re-writing a large part of it. I felt it was a worthwhile and steep learning curve and was greatly encouraged by the grade I received. I hoped that it would set the tone for the rest of the year.

METHODS OF ENQUIRY 1 (MOE1)

MOE1 was my first tentative step towards learning to be a researcher. The workshops were the platform from which to begin to develop an understanding of different methodological approaches and techniques for research.

The assignment was a research proposal for the topic I wanted to explore in MOE2. I was convinced that any high-quality research would need a robust plan which held together with regard to its epistemology, methodology, research questions, methods and ethics in order to be successful. I gained a significant appreciation of the need for ethical considerations and approval.

From the beginning of the course I had been searching for the appropriate methodology for my study. I was fascinated by the social constructionist approach in which there was no ‘one truth’ as supported by Wetherell, Taylor & Yates, (2001) because the process was interpretive, and the notion that members of a society actively created their own meaning from what they ‘do together’ as suggested by Burr (2003). At this point I was under the impression that I needed to pursue discourse analysis. However, I later discovered that I would not be interpreting the discourse so much as aiming to find themes within it and that thematic analysis would be useful here.

METHODS OF INQUIRY 2 (MOE2)

MOE2 was a natural progression from MOE1, allowing me to move towards my goal of making a practical contribution towards the improvement of WBA. I was able to apply my understanding of the theory and practice involved in educational research to a small-scale project looking at the richness of assessor texts. The study demonstrated that trainer attitudes to assessment feedback could be characterised as falling into three distinct types; Mentor, Examiner and Administrator.

At this stage I discovered that the method I wanted to use was thematic analysis (TA) rather than discourse analysis or grounded theory or content analysis or interpretive phenomenological analysis. I was not analysing for discourse or for narrative but for themes. TA of assessor texts was a novel approach in surgery. It was not well understood and in some areas of practice it was trusted less than quantitative approaches.

The advantages of doing the study as part of a taught course were the safe space it provided in which to trial and take risks while assuring readers of the validity of the process. At the outset I did not know what the outcome would be because the design allowed this to emerge as part of the process. However, I found seeing examples of how coding was done in the workshops and a similar study by my supervisor very enlightening.

MOE2 taught me new skills which gave me confidence to justify keeping my study small while retaining its meaningfulness. I felt it was in my power to go away and just do it. Coding and theming to find patterns in the data helped to make the familiar new. At the same time, I was aware of how subjective the process was and the importance of rigour and justification for research decisions. For these reasons I undertook to keep a research diary documenting all the steps I took and the reasons for them. The two main advantages were that they helped me learn from mistakes and construct a method for future use.

Reflexivity; examining myself as a researcher participant, was a completely new personal phenomenon for me. I became aware of the role that my own experiences, values and personal desires played in the decisions about the choice of research topic, methodology and approach to pattern analysis. I hoped that I added humility, practicality, efficiency and honesty to the process.

I was grateful for the opportunity to present my study to several internal curriculum committees. It met with a favourable response and constructive feedback. I was pleased that my Surgical Director suggested that I submit an abstract for an oral presentation to the Association of Surgeons of Great Britain

and Ireland (ASGBI) International Surgical Congress in Belfast in April 2016. I went on to present the outcomes there and at the Faculty of Surgical Trainers Annual Meeting in Edinburgh in October 2016 and published an article in *The Bulletin* of the Royal College of Surgeons of England: http://publishing.rcseng.ac.uk/doi/pdf/10.1308/rcsbull.2017.180.

I, therefore, felt that this research was recognised and helped to highlight the need for change in WBA. It also helped to change perceptions of me at work and in the process transformed my job description. In summary, MOE2 represented a challenge and became an achievement.

INSTITUTIONAL FOCUSED STUDY (IFS)

I had intended that MOE2 would be the first stage of an investigation which would develop through the IFS. MOE2 demonstrated that there was room for improving the way in which WBA was used. It allowed the study focus to shift to exploring ideas about how to aid better engagement of trainers with trainees as mentors with mentees, addressing trainee needs rather than curriculum processes. The IFS would, therefore, build upon MOE2 by proposing to determine a new method of feedback to be used with WBA by employing a mixed methods approach.

The study involved a group of trainee-trainer pairs participating in action research and required a good command of a range of methods new to me, including interviews, a workshop and a workplace trial. It entailed stepping outside familiar surroundings and discovering more about the surgical trainee and trainer view of reality. It provided an opportunity to consider different strands of thinking about how formative feedback between trainees and trainers could be better guided and more intuitively captured. The IFS culminated in the creation of a new tool

called the Feedback FREND. Although created in surgery, the Feedback FREND could be used by any teacher and learner in order to enrich verbal feedback through a more balanced and tailored dialogue. It promoted self-regulation in the learner and was a means of addressing the learner's specific concerns. The Feedback FREND tool, including a simple phone App, can now be found on the curriculum website www.iscp.ac.uk. It was presented at several surgical conferences including the Association of Surgeons in Training Conference in Bournemouth in April 2017 and at the induction to core surgical training in Iceland in February 2018. An article about the study was published in *The Bulletin* of the Royal College of Surgeons of England http://publishing.rcseng.ac.uk/doi/full/10.1308/rcsbull.2017.231.

THESIS

The thesis aimed to help address new standards for curricula from the General Medical Council. While MOE2 and the IFS were focused on improving the way existing competency-based assessment was employed, the thesis made a significant shift towards exploring the foundation for a brand new outcomes-based assessment before it was implemented in order to help align assessment theory with practice. Through the thesis, a new theory of Judgement in Action provided a perspective on how clinical supervisors approached the judgement of trainee performance and progression, promoting four types of capability.

I learned how to extend interviews into case studies with further use of thematic analysis. In both the IFS and thesis, I learned that in-depth interviews, albeit with small numbers of participants, coupled with robust and systematic thematic analysis could generate immense amounts of rich data and had the power to

uncover hidden relational and contextual factors that could not have been foreseen.

Conclusion

The progression across the courses and assignments has revolved around the development of an idea I had had at the beginning. Guided by tutors, who have had a significant impact on my assignments, I have been brought to the exciting realisation that I may be able to add something of my own to this area of research and practice. Studying while working full-time was very challenging at times and I have benefited in this endeavour from the assistance of many people set out in the acknowledgements. Along the way I have become more appreciative and tolerant of studies by others. I have learned how to be critical of even those who are leaders in their field by examining writings in a much more stringent manner, focusing on methods as much as results.

In summary, the effort I have made climbing to the summit of the EdD has given me a better view of the landscape of research in the area of education. I now need to move forward with confidence using the lessons I have learned to reach the goal of entering into the world of educational research.

Word count: 2085

Acknowledgements

Mr Anthony Bussey

First and foremost, I wish to thank my husband and critical friend without whose continual support this endeavour would not have been possible.

*

Mr Bill Allum and Mr Gareth Griffiths

I was fortunate to have received encouragement and support from successive Surgical Directors. They sponsored my EdD and made the research that underpinned it possible. They also provided invaluable sense-checking and criticism derived from an inside view of the practice of surgery.

*

Dr Gwyneth Hughes

I was lucky to have benefited from the expert advice given to me by my primary Academic Advisor. Her guidance kept me on track and her insightful perspectives and encouragement saw me through what at times felt like a very challenging process.

*

Professor Brian Cunningham

I am very grateful to my second Academic Advisor for his expert guidance and support.

*

Fellow students

I wish to thank my academic colleagues for their friendliness and for the helpful discussions. Sharing ideas and concepts helped me to better understand and contextualise.

*

Royal Colleges of Surgeons

This study would not have been possible without the approval of the Joint Committee on Surgical Training at the Royal Colleges of Surgeons.

*

Thesis participants

Finally, I wish to thank my study participants, the consultant surgeons who gave up their valuable time to provide me with helpful insights on their practice which greatly assisted this study.

Contents

List of figures

List of tables

Chapter 1: Introduction

> "In the varied topography of professional practice, there is a high, hard ground overlooking a swamp. On the higher ground, manageable problems lend themselves to solution through the application of research-based theory and technique. In the swampy lowland, messy, confusing problems defy technical solution. The irony of this situation is that the problems of the high ground tend to be relatively unimportant to individuals or society at large, however great their technical interest may be, while in the swamp lie the problems of greatest human concern."
>
> (Schön, 1990, p. 3)

The postgraduate medical training curricula set out by the medical royal colleges in the UK during the 2000s were highly desirable in terms of medico-cultural and political imperatives. They represented the 'high ground' that met the expectations of external bodies who sought accountability, transparency and responsive systems of training management. For those in training, however, the aspirations of the curriculum seemed remote from everyday work. They were situated in the 'swampy lowland' of medical practice, a heterogeneous, idiosyncratic culture, institutionalised around local practices and training norms. A gulf, therefore, existed between the high ground of those involved in planning the curriculum and the low ground of those using it. How best to bridge the gulf was a debate which formed the background to this study.

The development of medical training can be viewed as a journey which has moved through a number of stages from the personalised 'apprenticeship model' to the current de-personalised competency model. This study explored whether a new outcomes-based assessment, which purported to draw more closely on supervisor expert judgement and renew the recognition of surgical capabilities,

might bring about a revival of a more holistic view of practice and help to re-orientate assessment leading to a more complete picture of performance.

1.1 Study overview

The study is set out in eleven chapters exploring the application of judgement in surgical assessment. Chapter 1 describes the background to the study and outlines some of the shifts in educational theory that have led to a re-orientation towards expert judgement and the consequent impact on the structure of surgical training. It is followed in chapter 2 by a detailed account of the trajectory of the current competency regime and its proposed replacement with a new framework of Capabilities in Practice (referred to by the plural form of CiPs). CiPs represents a major paradigm shift in assessment by promoting the concepts of judgement, expertise and trust. Chapter 3 presents the research problem and the case for my enquiry, culminating in the research questions which divide the study into two parts, the first looking at the judgement practices of clinical supervisors and the second considering how far the new assessment corresponded with them. The literature is considered in Chapter 4 and is divided into what is known about judgement, expertise and trust, first through research studies conducted in general fields and second through studies based in clinical practice. The chapter concludes with a new definition drawn from the literature that clarifies the role clinical supervisors need to adopt when applying assessment judgement decisions. Chapter 5 details the research design, covering my semi-insider position and perspectives on the important ethical, theoretical, ontological, epistemological and methodological considerations needed to meet the study's objectives. It explains the choices that built on this qualitative, exploratory case

study approach that sought the accounts of clinical supervisors in their natural settings, recognising that they created knowledge and meaning through experience. Chapter 6 covers the data collection strategy, featuring interviews and a simulated CiPs assessment. It also explains the method of analysis which involved detailed and iterative coding and theming to interpret participant accounts, culminating in a list of the emergent themes. Chapter 7 presents the findings from stage 1 of the study. It describes the social influences that coloured the judgements participants made and the nature and quality of surgical decision-making. It also depicts the determinants of supervisor trust in trainees, how and why trust tended to be granted and what made trusting more likely. Chapter 8 develops a new judgement theory that illustrates the judgement pattern adopted by supervisors. This new theory, Judgement in Action, which draws on the theories of judgement and trust, incorporates four types of capability. In order to explore how Judgement in Action could be applied to assessment practice, chapter 9 presents the stage 2 analysis, looking at the effect of CiPs on participant judgements. It answers two research questions relating to how the participants' perspectives were altered by CiPs and how far CiPs resembled Judgement in Action. Chapter 10 covers the contribution that Judgement in Action makes to theory and practice, recommended next steps and the limitations of the study and is followed in chapter 11 by the overall summary and conclusion.

This study focused on the judgement practices of consultant surgeons who act in the role of clinical supervisor in surgical training, referred to below as *supervisors*. The learners, described as *trainees,* are postgraduate doctors who have chosen to specialise in the field of surgery. They work and learn under supervisors within surgical training programmes in NHS hospitals throughout the UK and Ireland and are not considered to be fully trained until they complete these programmes

(the period shown in mauve in table 1). The training process takes a minimum of ten years from the attainment of a medical degree to completion of specialty training, leading to certification, eligibility for entry on the Specialist Register of the General Medical Council (GMC) and appointment in consultant practice.

Medical School	Foundation training	Surgical training		Certification and eligibility for consultant practice.
5 years	2 years	Core 2 years	Specialty 6 years	

Table 1: The programme of training from medical school to certification

1.2 Background to the study

Surgery, as we know it today, is defined as the branch of medicine which manages patients with conditions which may be amenable to treatment involving operative intervention (curriculum proposal 2018). Historically, the practice of surgery was first identified, together with the law and theology, as one of three ‘classical’ or ‘learned’ professions. These professions indicated a vocation founded upon a distinctive body of specialised knowledge, providing service to others for compensation rather than business gain (Crook, 2008 pp. 11-12). Professionals were given the right to exclusive practice within rules and standards laid down by their authoritative associations. Their aim was to ensure high standards of practice, restrict entry and maintain ethical codes of conduct in order to demonstrate that they could monitor and control their own workforce (Lunt, 2008). A distinguishing factor between medicine and surgery was the emphasis for surgeons on practical skills gained through apprenticeships while physicians received a largely academic education.

The early system of surgical training by apprenticeship was designed to inculcate new members and produce within them a sense of community in a manner similar to the idiom ‘community of practice’ first used by theorists Jean Lave and Etienne Wenger (1991). Wenger summarised communities of practice as ‘groups of people who share a concern or a passion for something they do, and learn how to do it better as they interact regularly’. Communities of practice required three components; domain, community, and practice (Wenger, 1998, p. 1). Within surgery the sense of a community of practice allowed new members to transition from apprentice to master, a process defined by Lipsey as learning under ‘the tutelage of an impressive teacher’ in which an ‘unknown world of ideas and methods is freely opened to the apprentice’. This learning relationship between apprentice and tutor allowed ‘an assimilation not only of external method and intellectual concepts but also of attitudes, customs, history, sensibility’ (cited by Neighbour, 2004, p. 33).

The apprenticeship model, placing as it did the relationship between surgical trainee and supervisor at the heart of the learning process, continued to be strongly grounded in training up to the late 1990s. However, in educational terms, it had the propensity to be tacit with no standardisation or guidelines about the knowledge and skills to be taught or how long training should last (Franzese and Stringer, 2007). In the early 2000s, modern practice started to place pressures upon the apprenticeship model that were increasingly difficult for it to bear. The European Working Time Directive (Temple, 2010) for example, restricted the amount of time trainees could work in a week. The resulting shift work brought about a lack of continuity of both supervisors and training episodes and less exposure to cases and procedures. Increased numbers of trainees and increased

complexity of what had to be learned were also significant factors (Blank, 1982). As Eraut (1994) argued, a problem can arise when professions' preferred views of themselves do not conform to 'working realities' and, therefore, a new way of working had to be found to displace the profession's conception of how surgical apprenticeships should work. It became widely-held that formalisation was necessary and that it should be through a competence-based methodology.

The report *Unfinished Business* by Donaldson (2002), Chief Medical Officer for England, recommended that reform was needed to take account of poorly structured and unplanned medical training posts in which there was inadequate supervision, assessment, appraisal and opportunities for learning. In February 2003, the four UK Health Departments introduced the *Modernising Medical Careers* (MMC) initiative to oversee a major reform of postgraduate medical education in the UK (DH, 2014). MMC argued that examinations were biased towards the testing of knowledge and recommended that curricula took account of the skills and competencies required in practical employment. This led to the establishment of the Postgraduate Medical Education and Training Board (PMETB), merged from 2010 with the General Medical Council (GMC), to oversee the way in which the medical royal colleges structured curricula and delivered training. All curricula were expected to address clinical safety, expected levels of performance, standards of performance, transition points, assessments, patient expectations, equality and diversity legislation, strategic workforce issues and operational and professional objectives.

A tripartite commitment was made between the GMC, the medical royal colleges and the Schools of Surgery to improve UK training. The GMC was responsible

for setting the standards and approving postgraduate medical curricula, the medical royal colleges governed the content, structure and methodology, and the Schools of Surgery ensured its delivery. With these provisions in place, a new competency-based curriculum, the Intercollegiate Surgical Curriculum Programme (ISCP) was launched in 2007 (www.iscp.ac.uk).

Through an evolutionary expansion, my remit as Head of the ISCP, came to encompass the development of strategy and policy as well as the design of the trainee electronic portfolio, workplace-based assessment (WBA) methods and the learning agreement through which trainees and trainers provided evidence of training activities and learning. Working alongside surgeons, educationalists, psychometricians and web developers, I had particular responsibility for curriculum development. Promoting and justifying the curriculum were key goals, in terms of helping surgeon-educators understand the principles and strategies required by the curriculum. While I had faith in the concept of WBA, I was concerned that it was not well understood by those using it and wanted to participate in the discussions around increasing its effectiveness. My positionality, outlined in section 5.1, provided unique advantages, including access to the practitioners and a fresh perspective supported by the EdD, from which I hoped to offer insightful interpretations. My philosophy was that it was possible for the high and low grounds, expounded by Schön, to complement one another through a better understanding of what each required of the other.

The rationale for Competency Based Assessment

The term ‘competency’ first appeared in the early 1960s, leading to the development of a wide variety of different competence-based models. These

models had a variety of titles and purposes, and were considered variously as an approach to training, a form of assessment, and a model for curriculum design. In surgery, the term competence-based assessment (CBA) represented a model for curriculum design and the acceptance of the general definition of competence as ‘an ability to do something successfully or efficiently’ as set out in the Oxford English Dictionary (2018). The meaning of the term ‘competence’ was, however, widely debated. Parker (1984), believing the Oxford English Dictionary definition to be too ‘narrow’, offered a broader meaning of his own as ‘an orchestration of personal knowledge, skills, and attitudes relevant to the accomplishment of tasks.’ Ten Cate posited that the term ‘competency’ referred to the ‘skill’ itself while competence was the ability to perform that skill as an attribute of the performer (2013). In the clinical context, competence could be defined as the ability to make effective decisions or be competent to perform clinical tasks. Spady (cited in Tuxworth, 1989) defined competencies as indicators of successful performance and distinguished them from discrete cognitive, manual and social *capacities* such as motivation and sensitivity which served as enablers upon which competencies ultimately depended. Norris, (1991), drawing on Messick, defined competence as ‘what a person knows and can do under ideal circumstances’ (p. 333), suggesting that competence was about potential whereas performance was about the application of competence; what is actually done under existing circumstances. However, while these definitions were similar, and the term was used interchangeably, they were not synonymous. According to Brown, the modern approach to competence-based models evolved through five ‘generations’ of development (Brown, Patrick, Tate & Wright, 1994). The first generation closely conformed to the efficiency movement for quality control in mechanised areas of work like production lines; the second pertained

to education in order to promote the concept of mastery learning; the third applied psychology to the design of vocational educational programmes such as National Vocational Qualifications (NVQs); the fourth saw the development of behavioural objectives, while the fifth focused on outcomes derived from an analysis of work roles. According to Wolf (1995), CBA was first used in localised experiments in America in the 1970s. It was quickly adopted in the UK, Canada, Australia and the Netherlands as well as other countries. In the UK, government policy for NVQs in the 1980s endorsed its adoption by tying it to central government funding.

In surgery, CBA gained traction with an outcomes-derived analysis of patient-safe care. It was considered appropriate because of its binary ‘can’ or ‘cannot do’ nature which aligned with whether or not an individual surgeon could safely perform an operative procedure. The CBA approach extended and refined Miller's pyramid (1990) in which knowledge (knows), was at the lowest level of the pyramid followed by competence (knows how) and performance (shows how). Action (does) was at the peak of the pyramid and, with reference to surgery, focused on what occurred in practice rather than what happened in an artificial assessment situation. WBA targeted this highest level of the pyramid, collecting information about doctors' performance in their normal practice. Other common methods of assessment, such as examinations, targeted the lower levels of the pyramid. Underlying this distinction was the pragmatic suggestion that assessments of actual practice were a better reflection of routine performance than those done under test conditions (Norcini, 2003). Miller’s framework was highly influential among surgical educators and laid important groundwork for the acceptance of CBA. It set out the principle first described in Bloom’s taxonomy

(1971) that learners progressed from a basic knowledge of something through a series of levels from knowing to understanding then synthesising. The CBA model provided a basis for understanding and designing curriculum and assessment that were intended to measure actual professional practice (Rethans *et. al*., 2002).

CBA programmes were intended to address a number of perceived deficiencies in traditional teaching programmes which were criticised as making very general statements so that for learners, what had to be learned was not explicit. They placed considerable reliance on lecturing by instructors from textbooks and teaching in blocks, with a disproportionate emphasis on theory. Within each learner group, all learners spent the same amount of time on each unit, moving on together at the end of a block regardless of individual performance, meaning that assessment of learners tended to be norm referenced, promoting a uniform approach to progress and emphasising differences among students (Spady, 1977). These pedagogical programmes could be seen as deficient in terms of Blank's four propositions – i.e. they did not set out (i) *what was learned;* (ii) *how it was learned;* (iii) *when it was learned* and (iv) *if students learned* (1982).

In contrast, CBA programmes stated specific outcomes that should be learned on completion, for the first time giving learners a clear indication of what was expected of them. CBA's outcomes-based approach contrasted with the traditional form's stress on the 'process' involved in training. An 'outcomes' view recognised that there could be many means of achieving goals and, therefore, CBA programmes stated *what* was to be achieved but not *how*. Programmes provided supportive resources and tailored feedback (e.g. through formative

assessment), helping learners to master one element to an agreed standard before moving to the next at their own pace. The CBA philosophy reflected the belief that all learners did not reach the same level of competence for all activities. Similar to the notion of ipsative assessment, as advocated by Hughes (2014) and defined by the Oxford English Dictionary (2018) as 'designating or involving a measurement or scale calculated relative to a person's own performance or responses rather than those of others', CBA's individualisation also implied flexibility of time in terms of when, how long and how often learners were assessed. Gaining proficiency in one component at a time also meant that there could be recognition of prior learning within programmes, enabling learners to focus on gaining remaining outcomes (Brown, Patrick, Tate & Wright, 1994). To support flexible learning in this way, CBA programmes were intended to be modular. Modularity gave the learner the opportunity to choose and combine different relatively autonomous units of learning, building 'credits' in the system to achieve overall outcomes in different ways.

Another crucial difference between traditional systems and CBA was a realisation that the quality of teaching and the relationship between teacher and learner were at the root of success, supporting the notion that given the right kind of instruction and sufficient time, any learner could achieve mastery (Blank, 1982). CBA enabled the nature of the workplace environment in which learning was to take place to be a central focus of teaching (Brown, Patrick, Tate & Wright, op. cit.) and enabled the emphasis to be placed on practical 'learning from doing' *supported* rather than *led* by theory (Blank, op. cit.). The teacher followed a process of 'showing the learner' while the learner observed, then allowing the learner to assist and do parts while the teacher was available to intervene as

necessary. As the learner became practised at the basics s/he could proceed to more complex tasks. This reinforced the link with the concept of apprenticeships.

However, only programmes with all the above elements in place could be described as truly competence-based or classed as achieving 'high CBA' in terms of Gonczi's and Hager's three cumulative levels of CBA (cited in Brown, Patrick, Tate & Wright, op.cit., p.38). Most programmes claiming to be competence-based did not display all these ideals and could be described as achieving either 'low CBA' if they set out and assessed competencies in terms of standards or 'medium CBA' if they also included modular elements.

The problem with Competency Based Assessment

The application of CBA methodology in medicine, however, came to be criticised soon after its introduction. The literature supporting the CBA approach had been persuasive of the need to break down clinical practice into its constituent knowledge, skill and behaviour competencies (van der Vleuten & Schuwirth, 2005; Norcini & Burch, 2008; Reznick, 1993). The definition put forward by Wolf (1995) and widely adopted in America affirmed this point:

> "a form of assessment that is derived from a specification of a set of outcomes; that so clearly states both the outcomes-general and specific - that assessors, students and interested third parties can all make reasonably objective judgements with respect to student achievement or non-achievement of these outcomes; and that certifies student progress on the basis of demonstrated achievement of these outcomes."

This definition supported the assertion that there should be both macro-level (general) and micro-level (specific) outcomes. Specific outcomes at the micro-level were intended to be so clearly defined that both trained and untrained

observers should be able to understand unambiguously what was being assessed and what was being achieved (Blank 1982; Brown, Patrick, Tate & Wright, 1994), revealing CBA's roots in both criterion referencing and behavioural psychology. The broader theory of measurement by criterion referencing held that by objectifying criteria, surgical assessors could move closer to consensus. It legitimised the breaking down of any task into constituent elements in order to evidence the presence or non-presence of ability (Wolf, 1995). The behaviourist approach viewed tasks as 'data' which were directly observable even when removed from their context and meant that there was 'a tendency towards 'pseudo-quantification … the application of numbers to endeavours not readily amenable to quantitative analysis' (Collins, 1983, p. 176). An example of this could be seen when more importance was given to achieving numbers of WBAs and assigning scores and grades which could be summed instead of using them to give qualitative feedback, an approach that came to be known as 'tick-boxing'. With this view, CBA could be seen as 'a deterministic doctrine that was incompatible with the indeterminacy of everyday life situations' (Collins, op. cit., p. 178).

The risk inherent in this perspective for the design of curricula was that the emphasis on individual competencies could be at the expense of a more holistic picture of performance. It led to the tendency to describe competencies in exhaustive detail, leading to bulky, fragmented curricula that lost practical value for education as they became increasing disconnected with the real world (Ten Cate, Snell & Carraccio, 2010). Drawing on Stenhouse, Norris (1991) pointed out that there was a 'fundamental contradiction between the autonomy needed to act

in the face of change and situational uncertainty and the predictability inherent in the specification of outcomes' (p. 294).

Those favouring the ideals of CBA argued that it could be achieved in a humanistic setting if it was understood and implemented as originally intended (Ratcliff, 1984). There was only a risk when programmes compiled pre-determined 'laundry lists' of competencies and assigned them to learners. If, however, competencies were tailored to the individual learner, there was no determinacy about them. Norris (op. cit.) noted that the behaviourist constructs only needed to go as far as expressing what had to be learned in ways that made learning transparent, observable and measurable. In his view, the kind of human capability required could be inferred from performance. Parker (1984) noted that CBA did not attempt to sum up all the possible elements of behaviour that were important, just a sample of them in order to remove the 'mystery' from instruction.

However, a side effect of using unambiguous specific competencies was that it undermined the need for expert judgement in the situational context. This, according to Ten Cate and Scheele (2007), attracted criticism for attempting to de-professionalise teachers and assessors. They felt that mastery of competencies alone was not sufficient. What mattered was the use of judgement in deciding how and when to apply and combine competencies in context. Surgeons often stated, for example, that the most difficult judgement in surgery was deciding when *not* to act. Without this dimension one could master all individual competencies and still not be competent. Wolf (1995) argued that the inherent variability of the contexts in which competence was assessed and demonstrated meant that assessors had to constantly make major decisions

about how to take account of context when judging whether an observed action ‘fitted’ a defined criterion. Assessors, therefore, had to operate with a ‘complex, internalised, and holistic model’ of judgement rather than referencing a simple set of performance descriptors (p. 9). Wolf questioned whether it was likely, even in principle, that a combination of definitions and prior consensus would produce any very uniform behaviour and also whether the assumption of pre-existing ‘standards’ and shared understanding was reasonable at all. Meanwhile statistical studies had found that decisions made against individual criteria had far less power than global outcomes for judging performance (Regehr, MacRae, Reznick & Szalay, 1998).

According to Brown, the problems with CBA emanated from its top-down design. Academics and policy-makers did not involve the practitioners whose job it was to deal with day-to-day realities, instead imposing it on them in a way that was seen as an attempt to control the outcomes and processes of learning and undermine professionalism (Brown, Patrick, Tate & Wright, 1994). Winning (1994) argued that programmes generally operated through one of three different and conflicting paradigms which was informed by their philosophical assumptions. Her view was that CBA programmes tended to follow an *empirical-analytic paradigm*, taking a stance that reality had objective existence and believing that all knowledge could come from objectively proven facts through accountability and measurement. This view also saw programme content being determined and monitored by a ‘panel of experts’ which ensured existing positions of power remained. An alternative *situational interpretive paradigm* recognised that people and the world could not be separated because meanings grew from the way people interacted with the world. The implication for curriculum

design was that it should allow learning to be individualised, experiential, reflective and problem-solving in line with Dewey (1916) by favouring the 'by any means' achievement of goals. The last, *critical paradigm* would be concerned with uncovering the 'hidden agenda' in order to understand the fundamental interests, values, assumptions and implications for human and social action.

Criticisms in Surgery

In surgery, particular criticisms arose when attempting to interpret CBA frameworks through WBA (Swanwick 2005; Holzhausen *et al*., 2017). Assessment by checklists seemed to be an artificial construct because trainees and supervisors did not think in terms of competencies in daily practice and few supervisors were confident in their use (Regehr, MacRae, Reznick, & Szalay, 1998). As a result, the mastery of competencies failed to adequately prepare trainees for many of the realities of the context in which they were applied. Consequently, there was a building awareness that unless competencies were clearly linked to clinical care, they would be difficult to grasp (Ten Cate, 2005; Jones, Rosenberg, Gilhooly & Carraccio, 2011). Operative skill, for example, was a surgical competency which tended to be assessed through the undertaking of individual procedures such as performing a sternum puncture. Competency-based assessment might include the ability to follow a logical sequence, handle tissues well, control bleeding, use instruments appropriately and communicate clearly. These assessments were necessarily episodic in nature and were performed many times within single procedures. They determined on each occasion whether or not trainees had acquired the knowledge, skills and attitudes pertaining to the task. At the micro level, formative assessment and the surgical operative logbook (which documented all procedures undertaken and the degree of expertise) were useful for learning and auditing safety respectively. However,

disputes could arise as to how many operations trainees had to perform independently to be seen as competent and how far competence in one type of operation might determine competence in other types of operations. The chief drawback with reliance on binary assessments of this type, however, was the inevitable lack of comprehensiveness, meaning that a trainee could be technically proficient but the need for non-technical skills such as leadership could be missed. CBA had been adopted in surgery because of a desire to improve assessment through means that fostered objectivity, reliability and fairness. However, questions remained about whether formal assessment systems in postgraduate medical education supplanted the expert judgement of supervisors about their trainees leading to fragmented assessments of trainee capability. Schön's adoption of the metaphor of 'swampy lowlands' (1990), aptly described the environment within which training happened and its tendency to defy rational measures of assessment. It was widely held by the surgical community that the mastery of competencies alone was not sufficient to demonstrate that trainees were ready to progress (Rethans *et al*., 2002). What mattered more was the use of judgement about the context in which competencies were applied. However, basing assessment on clinicians' judgements of trainees would represent a shift in organisational mindset and require more evidence through research into the nature of judgement.

In summary, CBA replaced traditional programmes that tended to make vague statements about the content and goals of learning. These programmes were inflexible and heavily reliant on theory while lacking context. In contrast, CBA programmes were intended to be more transparent about what had to be learned, allowing for different progression rates and ways of achieving outcomes matched

to workplace requirements. They focused on the ‘does’ level of Miller’s pyramid, identifying performance as a product of competence. CBA for the first time gave learners a clear indication of what was expected of them by stating outcomes that they needed to have reached by completion. CBA allowed for flexibility in training, reflecting the philosophy that learners did not all progress at the same speed. However, there was a fundamental contradiction at the heart of the model. The intended consensus around the agreed criteria encouraged a focus on quantitative measures and accountability rather than qualitative judgements. Judgements about the quality of learning were being devalued, leading to the belief that the input of teachers and assessors was being de-professionalised. Ironically, a model designed to bring about an enhanced training regime was felt by many to be undermining its original intentions. Programme designers, therefore, needed to re-examine what mattered in day-to-day practice. They needed a pragmatic solution that could remove bureaucracy and re-value expert judgement of trainee progression. In surgical training there was a need to recognise that competence was more than the sum of separate competencies.

As covered in this chapter, CBA went some way towards meeting the objectives of an improved assessment regime but gave birth to new areas of concern. The next chapter moves on to depict the efforts that were made to design a new framework to counter the areas of weakness that arose from the way in which CBA was practically adopted by surgeons.

Chapter 2: The proposed response to the CBA problem

> "Education is not simply a technical business of well managed information processing, nor even simply a matter of applying 'learning theories' … It is a complex pursuit of fitting a culture to the needs of its members and their ways of knowing to the needs of the culture."
>
> (Bruner, cited in Swanwick, 2005, p. 43)

2.1 Entrustable Professional Activities

As patients, potential or actual, we all have an interest in the ability of surgical training to provide high quality surgeons. We expect the fully trained surgeon to have the requisite level of knowledge, skills and attitudes, the combination of which will define a competent surgeon. Competence models had been designed to ensure and monitor high standards of education and conduct, but CBA frameworks had tended to focus on assessing discrete knowledge and skills out of context and in a technical and fragmented way (Hauer *et. al.,* 2013; Fraser & Greenhalgh, 2001). Research conducted by Olle Ten Cate from the early 2000s promised to unite expert judgement with a CBA approach through the notion of the Entrustable Professional Activity (EPA). He proposed that EPAs could enable assessment to be set in context with a wider body of applied knowledge, skills and attitudes that constituted essential broad activities. Knowledge, skills and attitudes could be inferred from these activities over a number of occasions, demonstrating that trainees were not only able to perform procedures, but were also able to undertake ancillary tasks such as explaining to patients the rationale for an operative strategy and choice of operation, collaborating with auxiliary team members (e.g. nurses), organising the theatre set up and leading the team

briefing. He reasoned that if groups of competencies could be operationalised by linking them with professional activities, trainees and supervisors would be better able to understand precisely the extent to which trainees should successfully perform. Therefore, rather than taking a dualist view which saw CBA operating independently of holistic judgement, EPAs purported to enable each to complement the other to produce synergy and work together to produce an enhanced result.

Ten Cate described an EPA a ‘unit of professional practice that can be entrusted to a sufficiently competent learner or professional’ (Ten Cate *et al*., 2015, p. 983). EPAs allowed supervisors to make competency-based decisions on the degree of trust they could place in trainees and the consequent level of supervision trainees required in relation to their level of proficiency. At certain points in their education, medical trainees were expected to have attained sufficient competence to carry out clinical activities unsupervised. Entrustment through EPAs empowered supervisors to judge when trainees could assume different clinical responsibilities.

Ten Cate argued that competencies were so intertwined that assessing each of them separately made little sense and that the focus should instead be on inferring the presence of competencies from day-to-day activities (2006). EPAs were not an alternative to competencies, but a means of integrating them into clinical practice. While competencies were individual descriptors of tasks at the micro-level, EPAs integrated multiple tasks within essential broad activities at the macro-level, such as leading a ward round or out-patient clinic. EPA units of learning were situational and easier to assess as naturally occurring products

(Swanwick, 2005; Eraut, 1994), helping to address the tendency to assume that a competency could be applied equally in every situation. In this way they could be described as an immersive model (Bleakley 2006; Swanwick 2005).

Entrustment decisions were translated into four levels of supervision (level I, observe only; level II, execute with direct supervision; level III, execute with reactive supervision and level IV, supervise at a distance). An additional level V went beyond clinical competence to demonstrate that learners could also teach juniors. Supervisors would, therefore, use EPAs to judge the degree of oversight needed to ensure trainees performed safely and effectively. Entrustment was, therefore, closely linked to supervision and could be defined as:

The granting of permission to perform a function associated with patient safety.

Entrustment was based on how far supervisors felt trainees could be trusted to perform safely in certain situations or under particular circumstances. Trainee trustworthiness was associated with supervisor belief about what were important factors in relation to the task, the patient and the trainee. A trajectory towards incrementally greater entrustment represented trainee progress.

EPAs promised to restore supervisor expert judgement to assessment because supervisors were in a position to reliably judge trainees on their ability to accomplish professional activities which were grounded in daily practice. It was considered that in the clinical context, entrustment decisions were already a matter of daily clinical routine (Ten Cate, 2013).

In support of the theory behind EPAs, Swanwick argued that they drew more on ‘informal learning’, previously defined by Reznick as ‘characteristically collaborative, usually involving the manipulation of tools and leading to context-specific forms of knowledge and skills’ (2005, p. 39). The approach contrasted with both competence models and traditional ‘formal learning’ which tended to be decontextualized. Informal learning was generally agreed to be central to any form of learning that took place predominantly at work. Learning in response to imminent situations could be unplanned but intentional learning (Eraut, 2000).

Fraser and Greenhalgh (2001) argued that adults needed to know why they needed to learn something, and they learned best when the topic was of immediate value and relevance. They argued that while competence was about what individuals knew or were able to do in terms of knowledge, skills and attitudes, EPAs resulted in ‘capability’ which was more than competence. Capability involved the individual’s ‘ability to solve problems, to appraise the situation as a whole, prioritise issues, and then integrate and make sense of many different sources of data to arrive at a solution’ (p. 801). Capability also involved cognitive processing similar to creative behaviour. Fraser and Greenhalgh also argued that capability ‘could not be taught or passively assimilated as it was achieved through a transformation process in which existing competencies were adapted and tuned to new circumstances’ (p. 800). Swanwick (2005) argued that the learning process was one of acquisition rather than transmission and that the supervisor’s main tasks were of ensuring access to, as well as structuring, these experiences. Trainees, therefore, needed to be provided with continual opportunities to be stretched by the uniqueness of each context, and ‘capability’

was seen as the ability to access knowledge and make connections across different situations.

Importantly, EPAs reflected the use of global judgements made by supervisors as a result of observing trainee performance on a day-to-day basis. Global judgements had been shown to supply greater evidence of validity than checklists (Regehr, MacRae, Reznick & Szalay, 1998). The decision to trust drew on supervisors' judgements based on clinical experience, teaching experience and the holistic knowledge of trainees.

Ten Cate's work was highly influential in the arena of medical education because of his specialist knowledge of curriculum development and competency-based medical education. He had been closely involved with major curriculum reforms, education research, programme evaluation and educational development in the Netherlands. These credentials lent a great deal of weight to his propositions which steadily grew in response to the concerns over the way CBA was used. Entrustment was seen as a central concept for safe and effective health care, linked to the idea of being able to trust a qualified person to undertake an activity in a responsible manner. EPAs promised to help translate CBA to the world of medical practice.

2.2 From EPAs to Capabilities in Practice (CiPs)

Prompted by the work of Ten Cate and by misgivings about how CBA had been implemented, the GMC, as the regulator of postgraduate medical education, introduced new standards for curricula in 2017 which had to be met by 2020. The

critical change was a requirement that all medical colleges re-wrote their curricula in order to describe fewer, broader, higher-level learning outcomes that followed the EPA approach. Trainees were expected to be able to demonstrate a more rounded development rather than showing success in achieving multiple granular competencies, through an *outcomes-based* rather than *competency-based* curricula.

A ‘learning outcome’ was defined by the Academy of Medical Royal Colleges (AoMRC) as ‘an area of professional practice that the trainee is trusted to do unsupervised, once they have demonstrated the required competence’ (2017, p.9). Outcomes-based curricula focused instead on the ultimate aim of training; the competent surgical consultant. Trainees were, therefore, to be judged against the essential elements of everyday work that tended to be undertaken by consultant surgeons. The outcomes were considered to be broad capabilities within which there would be multiple competencies. There would also be implications for assessment which would need to focus on broader capabilities which were more meaningful to supervisors, allowing them to use their professional judgements on trainee progression.

The development of the surgical outcomes-based curriculum

The GMC defined a curriculum as ‘a statement of the aims and intended learning outcomes of an educational programme’ (2010, p.9). An aim of the surgical curriculum was that it would act as a guide for all involved in training. For trainees and supervisors for example, it should express what the end product of training should be, what should be learned and how it should be assessed. For NHS employers, the primary expectation might be clarity about the number and length

of training placements and supervisor role descriptions which would allow them to make available the necessary resources in terms of funding, faculty training and time in job plans. External course providers would be able to verify that their educational content aligned with the syllabus, while the public would be able to understand the differences between specialisms and the extent of expertise at certification. Therefore, a wide range of stakeholders would be affected by these changes.

The new regime would introduce a new and briefer design, promoting a set of common surgical high-level outcomes which could adapt to any specialty situation. The detailed syllabus would be obviated by an aspiration towards a more global combined judgement system and a new summative component in the assessment system.

The eleven specialty advisory committees of the royal surgical colleges composed of consultant surgeons and trainee representatives served as the expert writing group. They were led by an overarching surgical director, supported by educationalists and administrators and governed by a Management Committee. The key interest groups whose views were sought included trainee and trainer associations, as the main audience, as well as patient and employer representatives, to ensure that public expectations and service needs respectively were accommodated.

The development of Capabilities in Practice

A strategy employed by the curriculum developers, in order to save time, was to provide starting principles and draft frameworks upon which to consult. However,

a balance had to be struck between going too far towards pre-determining the curriculum and starting from nothing. Pre-determining curriculum content risked resemblance to the high-ground approach referred to by Schön (op. cit.) while a ground up approach might have been too time-consuming and lacking in focus.

With the delivery of curricula in mind, a first consideration was whether to adopt a new term for the proposed high-level outcomes. In different countries, the term EPA had different meanings and was being used in different ways. In Canada, for example, EPAs resembled one of the UK WBA methods (the Procedure-Based Assessment) while in the Netherlands, the EPA was likely to be further developed by its founders, leading to a divergence from its use in surgery. In view of this, the need for a new title seemed appropriate. Similar work being carried out by the Royal College of Physicians had led to the re-titling of the EPA as 'Capabilities in Practice' (2017). This term had been shown to have resonance in medicine. An advantage of using the same term as used by physicians was the authoritative weight it could lend for implementation across the country. It could also enable transference by trainees switching between these two large disciplines. This represented the rationale for choosing to refer to surgical outcomes as CiPs rather than EPAs. Therefore, in surgery, the high-level outcomes were named Capabilities in Practice or CiPs and were defined as: *the units of work essential for independent surgical consultant practice.* In surgery, nine CiPs were devised to cover the critical areas of surgical endeavour:

1) Managing an out-patient clinic

2) Managing the unselected emergency 'take'

3) Managing ward rounds and the on-going care of in-patients

4) Managing an operating list

5) Managing a multi-disciplinary meeting

6) Improving patient safety and delivering quality improvement

7) Carrying out and assessing the quality of clinical research

8) Acting as a supervising clinician

9) Working in the NHS

(CiPs 1, 3 and 4 were used in this study and are included in appendices 4-6).

Curriculum development often tended to work backwards from the end point of training, consistent with an outcomes approach. The agreed end product or standard was described as the ‘day-one consultant’ (represented in figure 1 by the large circle). Therefore, CiPs had to describe the job that needed to be done at the level of a newly appointed certified surgeon. From the point of view of job design theory in human resource management (Rush, 1975; Hackman & Oldham, 1976), any job could be broken down to a relatively few essential activities that the job-holder had to be capable of performing independently. If the job-holder could perform all the essential activities of the job without help then they could be said to be independent practitioners, or in the case of trainees, ready for certification and eligible to apply for consultant posts.

Each CiP (represented in the figure by the mauve circles) encompassed multiple tasks (represented by the pink circles) which demanded a range of skills (or underlying competencies) such as being able to interact with others, examine patients and take a patient history. It represented switching from vague terms such as assessing trainees as ‘good communicators’ to trusting trainees to ‘take consent’ which incorporated the ability to explain and listen to patients, answer their questions, gain rapport and write the medical record.

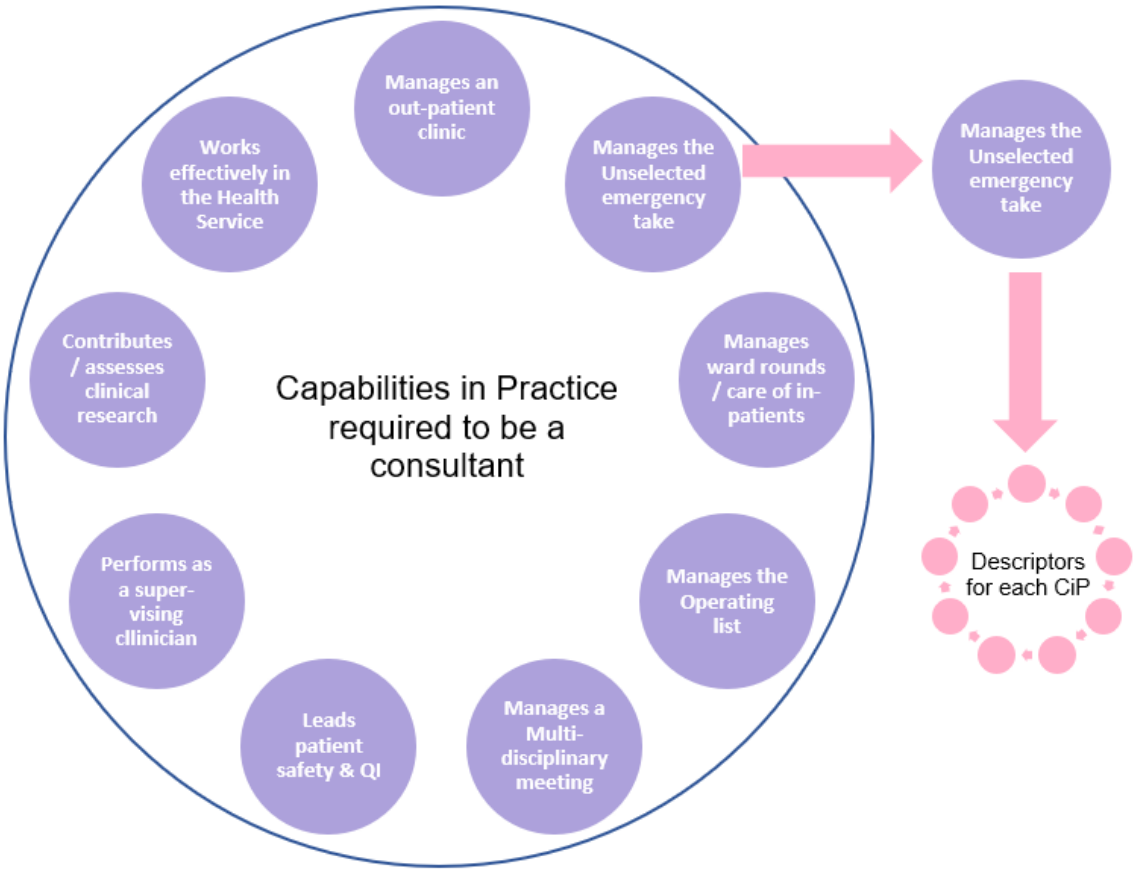

Figure 1: Capabilities in Practice as part of the surgeon's job role

Generic Professional Capabilities

Each of the nine CiPs was divided into two parts; clinical and professional (shown in figure 2), together expressing the make-up of a competent surgeon.

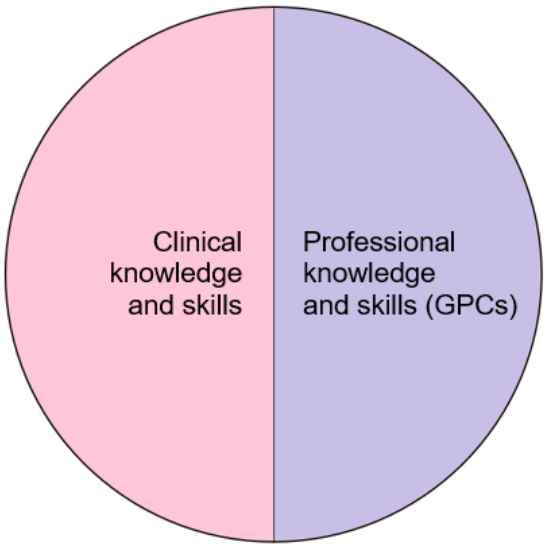

Figure 2: The two components of a CiP

The clinical component comprised the relevant tasks, knowledge and skills associated with broad clinical activity and was largely common across all eleven surgical specialties. The professional component had been developed by the GMC, in partnership with the AoMRC, and was called the Generic Professional Capabilities (GPC) framework detailing the essential values and behaviours that underpinned professional medical practice (see appendix 5 for abbreviated version used in the study). The GPCs reflected the concern that GMC fitness to practise cases related mostly to negative behaviours such as cases of malpractice by doctors involving poor professional behaviour e.g. dishonesty (GMC, 2016). The GPC framework, therefore, prioritised themes such as patient safety, quality improvement, health promotion, leadership, safeguarding of vulnerable groups and team-working. Consequently, the GMC and AoMRC mandated that the GPCs had to be a 'fundamental and integral part of all postgraduate training programmes' (op. cit., p. 3) with the aim of allowing early detection of issues and minimising the risk that any deficits would remain unidentified at the end of training. The nine overarching domains of the GPCs were as follows:

1. Professional knowledge
2. Professional skills
3. Professional values and behaviours
4. Health promotion and illness prevention
5. Leadership and team-working
6. Patient safety and quality improvement
7. Safeguarding vulnerable groups
8. Education and training
9. Research and scholarship

Plan for the introduction of the new CiP assessment

The GMC defined assessment as ‘a process of measuring a trainee’s knowledge, skills, judgement or professional behaviour against defined standards’ (2010, p. 8). It referred to programmes of assessment as an integrated set of assessment methods with different purposes linked to the curriculum through ‘a deliberate strategy and narrative for how and when learners will be assessed over time, particularly at critical progression points’ (op. cit. p. 9.). The aims of assessment in the surgical curriculum were to aid learning from reflection, demonstrate that learners had met the learning outcomes, and improve the assessment system.

The curriculum recognised two equally important but different purposes of assessment; formative and summative. Formative assessment, as defined by the GMC (2010), was primarily aimed at aiding trainee learning through constructive feedback that identified areas for development, and was managed through WBA. Formative assessment, however, contributed to summative assessment through the structured competence-based WBA forms held within trainee portfolios. Summative assessment was primarily aimed at determining a level of competence in order to permit progression of training or certification (op. cit.). The new CiPs assessment would be used for *both* purposes. CiPs would feature in the workplace throughout rotational placements in surgical units, offering trainees an additional formative assessment at the mid-point of each training placement and a new summative assessment at the end, leading to a judgement about whether they could progress to the next training level. The new assessment would call on Clinical Supervisors to fulfil the additional function of providing summative assessment (supervisor as examiner) while retaining formative assessment to support trainees (supervisor as teacher). A tension could result

from the challenge posed by combining these two different roles and responsibilities (GMC, 2011). Clinical Supervisors would be need to be able to distinguish between the marking of CiPs as a WBA and the marking of CiPs as a final assessment on the basis that the former contributed to the latter.

Hughes (2014) maintained that an ipsative approach could help support both formative and summative assessment. The notion of ipsative feedback was based on a comparison with the learner's previous performance linked to long-term progress and could have considerable motivational effects. An ipsative summative assessment would be based on progress towards criteria (e.g. a CiPs supervision level), rather than how far criteria had been met. An alternative to fixed criteria-referenced goals, therefore, might be to provide learners with achievable ipsative goals and opportunities to demonstrate how they had met these goals. It did not propose to abandon standards-based criteria altogether, but to provide individual criteria for each learner, which would enable realistic progression. This approach corresponded with the GMC's aim that outcomes-based curricula should allow trainees to progress at the own rates. The result might be a reduction in the number of trainees thought of as failing when perhaps only additional support was necessary. In surgical training, however, trainees would still be expected to pass waypoint examinations because of the critical nature of surgery and the need to ensure surgeons were patient-safe.

Supervision levels

A significant feature of CiPs was its rating scale which was influenced by the principles set out in Ten Cate's work on EPAs (2013) and the definition of trust set out on page 42, section 2.1 above as *the granting of permission to perform a*

function associated with patient safety. Supervisors would make decisions about trainee performance in terms of the level of supervision each trainee required in order to undertake whole activities (CiPs), instead of rating individual competencies, using the following scale:

Level I	Able to observe only
Level II	Able to act with direct supervision: a) supervisor present throughout b) supervisor present for part
Level III	Able to act with indirect supervision
Level IV	Able to act unsupervised
Level V	Demonstrates performance to a level well beyond that expected of a day-one consultant

Table 2: Supervision levels as a trust scale

The scale signified the degrees of nearness/absence of the supervisor with a gradual lessening of supervision until the natural step was to remove it. In a similar approach to that of EPAs, supervision levels were a way of demonstrating the degree of 'entrustment' to perform in a patient safe manner and, therefore, represented a trust spectrum. Level IV in all nine CiPs described a trainee who was ready to begin consultant practice. A trainee could demonstrate faster learning and excellence by either achieving level IV at an earlier stage of training or level V which was above the minimum threshold.

Through CiPs, supervisors would draw on their straightforward professional judgements of the broad clinical and professional areas of responsibility and their

experience of working with trainees in the context of their clinical practice. Using the scale, a supervisor might for example fully ‘entrust’ a trainee to carry out CiP 1 (managing an out-patient clinic) and CiP 3 (managing ward rounds and the on-going care of in-patients) by awarding a level IV. Their reasons might be that the trainee had demonstrated ability in organising clinics and ward rounds, knowledge of the key presenting conditions of the surgical unit, ability to undertake the out-patient and ward procedures and investigations and had worked well with the clinic and ward teams. They might, however, award a level II b) for CiP 2 (managing an unselected emergency take) and CiP 4 (managing an operating list) because the trainee was unable to discern priorities without help and was unable to carry out all the necessary operative interventions. Levels IV and V signified the unsupervised execution by a trainee.

Each CiP contained key descriptors to help trainees and supervisors recognise the level of knowledge, clinical skills and professional skills which had to be demonstrated for independent practice. These descriptors were intended to provide a language to describe trainee performance, requiring comments on only particularly excellent or poor performance. In situations where there was concern about a trainee’s performance there would be a programme of remediation including tailored or focused additional WBAs for trainees to undertake ahead of the final CiPs assessment. The assessment, therefore, provided a supervision level for each CiP with accompanying descriptors of strengths and weaknesses as well as free text. The CiPs assessment would occur at two points in every trainee placement. At the mid-point, it would be a formative assessment, providing trainee feedback. At the end of the placement it would be repeated as a summative assessment and would provide the key evidence for the ARCP (the

formal process under the Postgraduate Dean allowing trainees to progress to the next training year). The key differences between the current regime and the proposed new framework was that there would no longer be a long list of less than meaningful competencies that all trainees were expected to achieve in total at standard times. Instead, trainees would be assessed on a small number of whole activities which provided an individualised/personal picture of what they could be trusted to do under different supervision levels. CiPs would, therefore, provide supervisors with a different lens with which to see and judge trainees and represented a paradigm shift.

As covered in this chapter, the GMC's rationale that curricula should be re-focused around a limited number of high-level outcomes was proposed as a solution to the CBA problem, echoing the notion of the EPA. The response in surgery was to adapt and contextualise EPAs into nine CiPs to re-establish the judgement of supervisors. The next chapter sets out several areas of concern that needed to be addressed if CiPs were to be introduced successfully, ending with specific research questions.

Chapter 3: Research problem

> "Where is the wisdom we have lost in knowledge? Where is the knowledge we have lost in information?"
>
> (T.S. Eliot, n.d.)

The new outcome-based curriculum with its framework of CiPs set out to ensure that training promoted authentic supervisor judgements about trainee progression. Supervisors were thought to make judgement decisions about trainee performance every day based on a global sense of their abilities, albeit informal and unstructured. However, there were a number of problems that needed to be addressed. CiPs would significantly increase supervisors' responsibilities (AoMRC, 2017) but was highly theoretical and largely untested in the UK. The drivers behind complex entrustment decisions, which affected patient care on one hand and the learning and motivation of trainees on the other, were not clear. Exactly how supervisors made assessment decisions and what helped or hindered them was not well documented. It had been shown that expert judges tended to find it difficult to explain the logic behind their decisions (Bolger & Wright 1994). CiPs purported to be more in tune with real practice, but it represented a fundamental change in the way supervisors had been instructed to assess trainees to date. In the absence of studies in this area, the evidence of the success of CiPs would only become apparent after they had been introduced (Norris 1991). Surgical training also currently lacked the flexible timeframe for outcomes-based learning that would allow trainees to progress at varying rates (Watling *et. al*., 2016; Talbot, 2004). The current assessment regime demanded that all trainees reached specified waypoints at standard times and restrained the

development of those who were ahead or behind. An aim of the new outcomes-based curriculum was to allow trainees to progress at different rates. However, this would only be possible if supervisors could better discern and discriminate the abilities of trainees and recognise and value trainees who did not conform to the norm.

CiPs introduced more clinical accountability through expert judgement in which supervisors would need to assess risk (Damodaran, Shulruf & Jones, 2016). The study by Sterkenburg *et. al*., (2010) noted that at times there would be a need for learners to be allowed to do something on their own for the first time when their capability would be at the edge of their proven competence. The success of CiPs would rely on supervisors' knowledge of trainee practice together with a greater personal involvement in the outcome of an entrustment decision. Consequently, if judgements were not arrived at robustly, supervisors would be accountable too. Robust judgements relied on there being a training relationship of sufficient duration and good formative conversations, rather than the kind of variable quality of supervisor exchanges that were commonly fed into formative assessment (Bussey, 2016). Currently the difficulty in UK programmes was that close contact between trainee and supervisor tended to be of very short duration e.g. one episode a week over a 4-6-month placement, an issue also reflected in American programmes (Jones, Rosenberg, Gilhooly & Carraccio, 2011). CiPs were being introduced to the workplace very rapidly with a consequent risk for a culture which was relatively slow to adapt to changes in social practice (Swanwick, 2005). The supervisors, for whom CiPs were intended, were largely unaware of its existence and the implications its introduction would have for them. Some supervisors who

were still struggling to come to terms with the original regime, would need to be shown good evidence in order to be convinced of its legitimacy.

The literature on the use of EPAs (on which the concept of CiPs was based) was limited because although it had been strongly advocated, its adoption was in its infancy. However, because the GMC had endorsed its introduction into UK postgraduate medical curricula, it was reasonable to assume that its increased use would in time lead to more evidence of its performance. CiPs were likely to have a profound impact on trainee assessment and the duties of supervisors. The supervisors who would use CiPs, were not currently accustomed to making explicit summative judgements and their opinions were often ad hoc and verbal with little written down. They would need to accept the increased levels of accountability and the attendant risk of setting supervision levels without *knowing* whether trainees were ready to undertake patient-critical work.

The new regime would require a great deal of resource for its implementation in terms of training and induction. Trainees and supervisors would require protected time and a conducive working environment in which mutual respect between trainee and supervisor was essential. Supervision would need to be continuously available and easily accessible with feedback to supervisors on their performance. Finally, the professional behaviour of trainees, an area about which supervisors tended to be less clear (Bussey, 2017), would also need to be robustly assessed before progressive independence could be granted. The outcomes-based principles of CiPs would, therefore, challenge established and time-honoured practices that might be resistant to change.

No studies to date had been conducted into the judgement process surgical supervisors used to assess the ability of surgical trainees. Questions also remained about the impact the working environment had on this cognitive process. It was important to understand these factors in order to utilise expert judgement in assessment, check the fit of any new assessment with working practices and gauge supervisor readiness for increased levels of accountability. These problem areas fell into three categories summarised below.

1. *In relation to supervisor judgement:*

- How did supervisors make decisions about trainee progression?
- How conducive was the working environment to helping trainer judgements?
- How were trainees allowed to perform activities on their own for the first time?
- What knowledge of trainee practice did supervisors gain from the training relationship?
- What personal involvement in trainee progress was felt by supervisors?

2. *In relation to the new assessment:*

- How far did the new assessment correspond with the way supervisors currently made judgements about trainee progression?
- How far did the new assessment provide discernment about differing trainee abilities and allow for trainees who did not conform to the norm?

3. *In relation to implementation:*

- Was there evidence that supervisors would accept the increased levels of accountability proposed by the new assessment?
- What modifications were likely to be needed for implementation?

While this study sought to address the questions under the first two categories above, the third fell outside of the remit of this study.

3.1 Research questions

This study would, therefore, explore how supervisors currently judged trainees, aiming to unlock their thinking and the factors and issues that affected their current judgements, emphasising any differences in supervisor judgement processes. It would explore whether and how far, through the lens of three CiPs, judgements about trainees could be made more easily or more robustly, whether trainee progress could be more easily stratified and what problems and solutions might arise. The table below sets out the inquiry in terms of the broad questions, sub-questions and methods to be used.

<table>
<tr><td colspan="4">Overall research question: How do Clinical Supervisors make assessment judgements?</td></tr>
<tr><td colspan="2">Broad Questions</td><td>Sub-questions</td><td>Methods</td></tr>
<tr><td rowspan="3">Stage 1: Pre-exercise</td><td rowspan="3">1. How do supervisors identify when trainees are ready for increased levels of independent practice?</td><td>a) What factors influence supervisors, judgements about the ability of trainees to practise independently?</td><td rowspan="3">• Interview Part 1 (semi-structured / individual)</td></tr>
<tr><td>b) What cognitive processes do supervisors use to judge trainees?</td></tr>
<tr><td>c) What patterns of expert judgement occur in practice?</td></tr>
<tr><td rowspan="2">Stage 2: Simulation exercise</td><td colspan="2">2a) To what extent do supervisors’ perspectives of trainees alter through the use of CiPs?</td><td rowspan="2">• Simulation exercise to complete 3 CiPs
• Interview Part 2 (semi-structured / individual)
• Analysis of individual cases and cross-case synthesis</td></tr>
<tr><td colspan="2">2b) To what extent do supervisor judgements resemble the CiP approach?</td></tr>
</table>

Table 3: Research questions

As covered in this chapter, while CiPs purported to re-establish authentic clinical supervisory judgement to assessment, a fundamental change to the assessment regime, the proposal was currently theoretical and untested. How supervisors made judgements was not well documented and supervisors themselves found it difficult to vocalise their judgement decisions. In order to help convince trainees and supervisors of the legitimacy of the new approach, there needed to be evidence through enquiry (as set out in table 3) which in turn needed to be informed by the current literature and thinking. The next chapter explores the literature that helped guide the study design and analysis.

Chapter 4: The literature relating to expertise, judgement and trust

> "Experts, at least in areas involving non-empirical knowledge such as morality and mathematics, had, in another life, Plato said, learned the principles involved, but they had forgotten them."
>
> (Dreyfus & Dreyfus, 2005, p. 781)

The Dreyfuses argued that the rules governing the application of expertise resided in the expert's mind, whether or not s/he was conscious of it. They suggested that when learning something new, an individual had to think very hard about the steps involved but when the required skill was fully realised, that knowledge no longer needed conscious attention and the memory of its component parts could be lost.

The literature linking expertise, trust and judgement was underpinned by the notion that judgements were made by experts who belonged to trusted professions. Traditionally, professions were denoted as occupations involving prolonged training and formal qualification with high standards of education and codes of conduct. They allowed their members to be placed in positions of trust by their clients, society and the state to exercise judgement on the basis of their unique specialised knowledge – a kind of 'social contract' (Lunt, 2008). However, the autonomy enjoyed by professionals came to be challenged and undermined by reforms in the early 1980s which introduced measures such as regulating bodies, performance measurements, targets and outcomes-based evaluation. A re-orientation of the surgical curriculum towards the fostering of subjective expert

trust judgements made by supervisors rather than the objective, competency-based approach, therefore, represented a fundamentally different course. Any study seeking to validate this re-positioning needed to focus on the literature around the established and current thinking on expertise, judgement and trust in order to enlighten but not pre-determine the study outcomes.

The literature in this chapter is divided into six sections, covering in turn the general and clinical literature around these three areas. It concludes with a new definition drawn from the literature which might assist curriculum developers in framing and communicating to supervisors the nature and responsibilities of the role that they have undertaken.

4.1 General literature on the nature of expertise

Beyond establishing that experts were those with unique knowledge and skills that were highly contextualised to particular domains of practice, the study of expertise in a wide range of fields has shown that expertise was difficult to define. Bolger and Wright (1994), wrote that it was normally ‘accepted that we know an expert when we see one’ (p. 2) but quoted Johnson’s broad outline as an example of the characteristics many authors believed expert decision-makers should possess:

> "An expert is a person who, because of training and experience, is able to do things the rest of us cannot; experts are not only proficient but smooth and efficient in the actions they take. Experts know a great many things and have tricks and caveats for applying what they know to problems and tasks; they are also good at plowing through irrelevant information in order to get at the basic issues and they are good at recognising problems they face as instances of types with which they are familiar.

> Underlying the behaviour of experts is the body of operative knowledge we have termed expertise."
>
> (Johnson, cited in Bolger & Wright, 1994, p. 2)

The major impediments to uncovering the general features of expertise were its complexity and domain-specific nature. Additionally, experts in different fields have themselves tended to find it difficult to articulate the principles underlying their expertise and cognitive processes (Dreyfus & Dreyfus, 2005; Feigenbaum & McCorduck, 1983). Expertise, therefore, appeared to be a hidden rather than an overt aptitude as it tended to be the result of experts' experience of searching for relationships between the subject matter and the reality to which it referred. This type of application-orientated learning process resulted in a deep knowledge of the concepts and principles of a subject area (Litzinger, Lattuca, Hadgraft & Newstetter, 2011) but with an opacity that meant that experts did not always know exactly what they knew about their domain (Dreyfus & Dreyfus, 2005). The search for a better understanding of expertise, therefore, led to a number of observations and categorisations about what made experts expert.

Cognitive science sought to understand the mechanisms arbitrating expertise in various domains, including chess, sport and medicine, and debated whether a common set of theoretical principles could encompass them (Ericsson & Smith, 1991). Ericsson and Smith hypothesised that there were stable personal characteristics that underpinned expertise and that these could be categorised as either inherited (e.g. personality) or acquired (e.g. skills) and as either general (e.g. sociability) or specific (e.g. suturing) as summarised in table 4. They found that expertise was predominantly acquired, and posited that it could be measured

by designing experimental tasks to elicit the critical aspects of expert performance.

	Inherited	**Acquired**
General	Personality / Sociability	Skills
Specific	Steady hands	Suturing

Table 4: Types of stable personal characteristics

Other ways of categorising expertise involved focusing on the attributes experts were thought to display. These included *personal characteristics* such as high stress tolerance (Shanteau, 1988); *competence* such as searching and filtering ability (Johnson, 1988); *mental processes* such as perceptual and sensory awareness; *cognition* such as problem-solving (Anderson & Lopes, 1974) and *morality* such as ethical conduct (Barber, 1963). The Dreyfus brothers asserted that intuitive judgement was the hallmark of expertise (2005).

Shanteau (1988) went on to make a distinction between *substantive experts* (whose skills lay in analysing large bodies of data) and *assessment experts* (whose skills lay in making judgements despite uncertainty). He found that substantive experts tended to have (i) *cognitive expertise* (possessing unique problem-solving abilities enabling them to discover relations not found by others e.g. accountants); (ii) *knowledge expertise* (making decisions based on a large amount of information e.g. academics) and (iii) *advice expertise* (providing information to others without necessarily acting on it themselves e.g. counsellors). In contrast, assessment experts tended to have (i) *perceptual expertise* (enabling them to perceive differences in people not apparent to others e.g. judges); (ii) *diagnostic expertise* (making decisions despite considerable

uncertainty and limited information e.g. medical doctors) and (iii) *action expertise* (carrying out their decisions e.g. airline pilots). Experts were also thought to share psychological characteristics such as perceptiveness, pattern-recognition, communication skills, self-confidence and creativity under stress. They were able to convince others of their expertise through good social skills and were more resilient than non-experts.

Drawing on the work of Bloom, Ericsson, Prietula and Cokely (2007) affirmed that there was no correlation between IQ (the intelligence quotient designed to assess human intelligence) and expert performance in fields such as chess, music, sports, and medicine. Bloom had explored superior performance and revealed that the amount and quality of practice were key factors in the development of expertise. He showed that experts learned their skills from teaching, practice and experience rather than through natural possession (i.e. experts were made not born). The journey to expertise required struggle, sacrifice, and honest self-assessment. It was only deliberate practice that developed expertise, entailing considerable, specific, and sustained efforts to develop in a domain that originally could not be done well or even at all. These theories coincided with views emerging from studies in cognitive science and there was consensus suggesting that expertise was a rare skill that developed only after 'much instruction, practise, and experience' (Camerer & Johnson,1997, p. 195).

Although expertise was defined by different writers in different ways across different fields, their views coalesced around a notion that expertise could be developed through deliberate practice and had merit in providing a service in a contextualised field.

4.2 Literature on clinical expertise

Most studies on expertise in medicine have tended to concentrate on purely clinical rather than clinical-supervisory expertise. However, a useful general model postulated by Dreyfus and Dreyfus (2005), theorised that expertise developed through five stages from *novice* through *advanced beginner, competent performer and proficient performer* to *expert*:

1. The *novice* followed context-free rules, a rigid approach which lacked the flexibility to adapt to the real world.

2. The *advanced beginner* developed 'instructional maxims' devised from limited experience coupled with the objectively defined non-situational features of the novice.

3. The *competent performer* devised rules and reasoning procedures to arrive at an appropriate perspective in order to manage the information overload created by increasing experience. Deliberate choices of this kind were personal investments which tended to give rise to emotions related to success or failure unknown to the first two stages.

4. The *proficient performer* could discriminate among a variety of situations, see goals and distinguish salient aspects, but still had to fall back on detached rule and maxim-following.

5. The *expert* not only understood what needed to be achieved because of a vast repertoire of situational discriminations, but also comprehended immediately how to achieve the goal. Further, an expert would try to protect against a narrow outlook that might disregard possible outcomes by being sympathetic to alternative views sometimes through reflection and sometimes by consulting others.

The Dreyfuses' model suggested that the cognition of experts was more sophisticated than that of novices and that experts had better insight about the accuracy of their own predictions. As a result, they were likely to be more realistic in their actions. Shanteau's distinction between 'novice' and 'naive' was also important (1988). Novices (such as surgical trainees) might still have considerable and appropriate knowledge and experience while those who were 'naïve' (e.g. the lay public) were beginners in every sense. However, the linearity of the Dreyfuses' model did not allow for the possibility that those at any one stage could display the features of the other stages in particular situations, something which was very pertinent to the field of surgical training.

To date, the outcomes of studies have provided both categorisations of expertise and theories of how expertise might develop but little is in the area of clinical expertise applied to supervision.

4.3 General literature on judgement

According to Coles, the nature of professional judgement 'is revealed when society asks experts to undertake certain tasks and perform certain roles that

others cannot or will not do' (Coles, 2002, p.1). In surgical practice for example, consultant supervisors were responsible for ensuring that trainees were patient-safe, a role that required not only clinical judgement but also an understanding of educational assessment and a close working relationship with trainees. According to Bolger and Wright (1994), however, crediting expertise to individuals in particular roles did not necessarily depend on their knowledge and skills.

According to Ericsson (2004), there were three tests of expert judgement. First, judgements had to lead to performance that was consistently superior to that of non-expert peers. Second, judgements had to produce concrete results. Finally, the results of the judgement should be amenable to being replicated and measured. In terms of surgical supervision for example, supervisors might demonstrate passing these three tests if they used formative assessment to precisely identify trainee development needs and show improved trainee performance.

However, early studies on judgement had tended to view human judgement performance as biased and sub-optimal (Shanteau, 1988). Bolger and Wright (1994), wrote that the theory of conservatism (unwillingness to change opinion following updated information) was likely among those who judged. Human decision-making was seen by writers such as these as relying heavily on judgemental heuristics (information-processing rules or mental rules of thumb) as a shortcut to making decisions or judgements. Heuristics were developed by people as a means of coping with information overload, but they often led to biases or judgemental errors relative to normative standards (Tversky & Kahneman, 1974). In assessment, biases also generally occurred because

experts based their decisions on non-relevant criteria such as assessors judging student performance based on student background and culture. In a similar way, expert judges were also postulated as thinking in terms of complicated 'configural' rules when making predictions. Configural rules were defined as self-determined internal plausible narratives or verbal protocols (Camerer & Johnson, 1997). They were also defined as rules in which the weight of a given attribute depended on the level of other attributes (Ganzach, 1997). These rules could emerge naturally from past 'special' cases and were reinterpreted and refined over time. However, they were often applied inappropriately because while they were induced under rare conditions, they tended to be over-generalised (Tversky & Kahneman, 1974). They were also thought to be frequently inaccurate because of errors in the implicit theories that underlay them. Experts tended to treat an episode in which an informant provided information as potentially connected to the stable properties of informants and stable properties of the world. They assumed informant actions were caused by their unobservable psychological properties e.g. knowledge and intent. Informants were not thought of as idiosyncratic but as members of social categories to which specific inferences could be generalisable (e.g. generalising that trainees from a particular university would be better learners). Configural rules were thought to persist because they were easier. Feedback to decision-makers about the outcomes of their decisions was crucial to their learning, otherwise they tended to search instinctively for evidence that confirmed prior theories. Views such as these led to the development of decision aids similar to competency-based checklists to improve or replace judgements. They also led to the introduction of rational and predictive systems. Studies showed that what experts knew did not always enable them to out-predict simple statistical programmes. In a broadcast featuring Dr Zoe Williams, a GP, for

example, an artificial intelligence programme was able to diagnose a complex patient problem by asking questions in exactly the same way as she would have done (Williams, Leonard & Crabtree, 2017). However, the main advantage of using predictive or knowledge-based systems as ‘artificial experts’ was perceived to be in the ability to overcome the cognitive limitations of human experts (Shanteau, 1988; Bolger & Wright 1994). Artificial systems contained rules that experts were considered to use. They strove to match but not exceed the performance of the experts they represented. This viewpoint aligned closely with a behavioural science perspective which attempted to reach conclusions through rigorous formulations based on observation and experimentation.

The theories favouring artificial systems were, however, countered by the arguments of a number of authors. Harlen and James (1997) cautioned that any system that diminished critical thinking and problem-solving could have the effect of reducing what was assessed to what could be readily and reliably marked. Coles (2002) posited that any task that circumvented expert judgement left merely technical work. Coles also emphasised that human judgement processes were not as concerned about coming to the ‘right’ answer (which may not often exist in an absolute way) as reasoning what was ‘best’ for the situation. Educational assessment, as noted by Harlen and James, involved judgement about what was relevant evidence for a particular purpose and how to communicate it for an intended use. Dreyfus and Dreyfus (2005) concurred with these views, writing that judgement could not be reliably captured in rule-based expert systems. Bloxham’s 2009 study, exploring assessment and moderation in higher education, concluded that while there was a need to maintain the confidence of key stakeholders in making judgements, it should be done, not

through focusing on unattainable reliability and accuracy, but by emphasising professional judgement moderated by the internal and external HE community.

Camerer and Johnson (1997) researched decision-making across a wide range of professions including psychology, medicine, academia and accounting and argued that experts frequently demonstrated a number of traits that impacted on their choice of judgement strategy. Individuals who were considered as able to make expert decisions were situationally aware of rare factors that could affect a particular set of circumstances. An example of this might include a supervisor discerning which trainee might learn best from a particular set of circumstances through knowledge of both the trainee and situational factors. Experts knew of rare sets of factors, also known as 'broken-leg cues' which were highly diagnostic rare predictions by those with privileged information:

> "A clinician is trying to predict whether or not Professor A will go to the movies on a given night. A regression model predicts that the professor will go, but the clinician knows that the professor recently broke his leg. The cue 'broken leg' probably will get no weight in a regression model of past cases, because broken legs are rare. But the clinician can confidently predict that the professor will not go to the movies."
>
> (Camerer & Johnson, 1997, p. 205).

Experts also knew more than non-experts and also used their knowledge to guide searches for limited subsets of information which differed with each case. This strategy allowed more efficient encoding of task-specific information (Shanteau, 1988). Experts tended to seek group feedback, were willing to make adjustments to their views and divided tasks, helping them overcome the effects of cognitive limitations.

Experts in different fields tended to develop their own informal decision aids, consistent with evidence-based practice, pushing the practitioner to improve the quality of their decisions by systematically reviewing information from rigorous data-gathering efforts instead of relying on customary practice. This approach also allowed them to avoid the biasing effects of heuristics. However, it was noted that evidence alone was never sufficient to make important decisions such as those related to clinical practice, reasoning was also required (Coles, 2002).

Experts often used a dual strategy of first making a rough estimate and then conducting a more careful analysis. They simplified large problems by breaking them down into smaller more manageable parts, progressively finding solutions to the parts until the whole problem was resolved. However, they were also selective in picking decision problems, while novices were more likely to adopt one of two extremes; either taking on all decisions in order to be perceived as 'decisive' or avoiding making any decisions at all (Dino, Shanteau, Binkley & Spenser, 1984). Experts were revealed as having a sense of what was relevant and irrelevant when making decisions and appreciated that decisions that came close were good enough.

Overall, the exercise of professional judgement was thought to perform an important public role. The literature suggested that to be considered expert, judgement had to be delivered as part of a strategy which might overcome biases and heuristics. While artificial systems were thought to be capable of mimicking the judgement of experts, they tended to be limited to problems which lent themselves to a technical and rational solution. Human judgement, in contrast, was thought to be essential for problems that were not readily and reliably

encapsulated. They involved the use of reasoning and pragmatism in order to make more reliable decisions.

4.4 Literature on clinical judgement

Most studies on judgement in medicine have focused on clinical judgement applied to patient problems rather than on the judgement of trainees by supervisors. Although both types of judgement had elements in common such as deep knowledge of the subject and the complexity of the decision, affecting patient safety, studies focusing on clinical judgement emphasised diagnosis, therapy, prognosis and choice of tests and tended to be quantitative in nature. Social judgement theory (Wigton, 1996) for example tended to analyse the weighting of clinical cues from paper-based vignettes and often found wide variation among clinician judgements. The outcome of such studies suggested that clinicians did not make decisions in the way they thought they did, that there was great variation in the use of information and that strategies often differed from what was recommended by experts or found in textbooks. These facts were an important challenge to those trying to understand clinical judgement. However, these quantitative approaches had not been shown to be practical for applying to judgement related to clinical supervision. The high-stakes nature of clinical assessment led to the desire to enhance the political legitimacy of judgement by measuring its objectivity, reliability and fairness. It was suggested that predictive accuracy was an appropriate measure of expert judgement (Camerer & Johnson, 1997), consistent with the tendency to rely on the judgements of those who had been right in the past (Harlen & James, 1997) e.g. checking that supervisor judgements were borne out by actual healthcare outcomes. It often led to expert

judgement being compared with the accuracy of statistical systems in performing medico-judgement tasks.

Coles wrote about the nature of professional judgements in clinical practice, drawing attention to the distinction made by the ancient Greeks between two forms of human action; poesis and praxis. Poesis referred to actions where the means and ends were prescribed and required technical knowledge (techne). Praxis, on the other hand, referred to decisions about both the means and the ends and required phronesis (practical wisdom), believed by Aristotle to encompass the work of caring professionals. Coles went on to assert that evidence-based practice could deal quite well with data but might not inculcate practical wisdom (2002). He referred to Carr's assertion that someone who lacked phronesis 'may be technically accountable but can never be morally answerable' (p. 4). This approach had merit in surgery in which the work was characterised by uncertainty, complexity and ethical responsibility for those who were vulnerable, and which sometimes lacked clear resolution (or the swampiness referred to by Schön (1990). Coles' research with Fish (1998) was inspired by Grundy (1987) and the writings of Habermas (1972) and described a four-part typology of the kinds of judgements made by healthcare professionals. They labelled them as: *intuitive*, *strategic*, *reflective*, and *deliberative*.

1. *Intuitive judgement* was concerned with urgent practical problems needing an immediate response, answering the question '*What do I do?*' With no apparent reasoning involved, it was similar to Tripp's *practical judgement*, (2011).

2. *Strategic judgement* answered a procedural question most often associated with the passive following of protocols for relatively routine and clear-cut problems, it answered the question '*What might I do?*'.

3. *Reflective judgement* targeted situations where there were uncertainties which required a capacity for deeper thought, answering the question *"What could I do?"* This was closer to what they defined as professional judgement and was similar to the term used by Tripp (2011).

4. *Deliberative judgement* was an extension of reflection and of a higher order akin to the notion of professional judgement. Like phronesis, it was characterised by the moral question*, "What ought I do?"* The moral dimension - doing good for another person - was a type of professional judgement requiring high levels of support and structure but as little direction as possible and required an examination not just of a practitioner's actions (including the actions they chose *not* to take) but also an examination of the thinking that lay behind them. Deliberation was needed to develop expertise and was the result of considerable practice, reflection and analysis.

Deliberation referred to performance with the intention of improvement (Litzinger, Lattuca, Hadgraft & Newstetter, 2011). Dreyfus and Dreyfus (2005) considered that when deliberation was undertaken, it was for the purpose of improving expert intuition and to take responsibility for mistakes when they occurred, rather than trying to prevent them by fool proof rules. In line with this, experience alone was less important than learning from experience (Shanteau, 1988). These four kinds of judgements formed a hierarchy in respect of both higher-order thinking and the

types of judgement made by professionals. However, Coles and Fish did not suggest that novices began by making intuitive judgements and then developed the capability to make strategic, reflective and deliberative judgements sequentially. They argued instead that these types of judgements were used at different times as appropriate. Coles and Fish, therefore, fundamentally challenged the work of writers such as Dreyfus and Dreyfus (op. cit.) who delineated progression from novice to expert in a linear way.

Croskerry, Petrie, Reilly and Tait (2014) agreed that clinical decision-making was a complex process with many independent and interdependent elements. They distinguished two broad types of decision-making process in clinical practice to help manage complexity. *Intuitive* decisions were autonomous and typically fast, requiring minimal cognitive resource. They were largely based on pattern recognition which allowed considerable time-economy by matching existing patterns to particular decisions and actions. *Analytical* decisions, on the other hand, tended to be slow and deliberate, demanding conscious effort. These two processes were not differentiated by speed alone but rather the degree to which the problems could be characterised e.g. simple or complex. Decision-making tended to involve both processes, but different situations required different approaches and it was, therefore, important for practitioners to understand when and how to employ the most appropriate mode.

In summary, views about the judgement of human experts have varied between those who suggested that it tended to rely heavily on heuristics as a means of coping with information overload and led to assessment biases (Tversky & Kahneman, 1974) and those who advocated that it was a rare skill that developed

only after much instruction, practise and experience (Camerer & Johnson,1997; Coles & Fish, 1998; Coles, 2002). Currently assessment systems tended to undermine experts by introducing artificial expert systems such as predictive algorithms and competence-based rules. A useful alternative to these extremes, however, was to encourage the judgement of human experts to be assisted by evidence and decision aids that guided and formalised their decision-making process while still drawing on human knowledge, experience and social skills. The concept of deliberative judgement appeared to closely correspond with the type of judgement required of practising clinicians. Deliberative judgement had a moral and ethical nature and involved dealing with conflict (e.g. between safety and training) and uncertainty (e.g. acting with limited information). It involved acting independently and reflecting on practice for the purpose of improvement.

4.5 General literature on trust

Schön characterised the nature of professional work as uncertain and unpredictable. His adoption of the metaphor of ‘swampy lowlands’ (1990) corresponded with the idea of trust within a ‘risk society’ described by Beck (1992). Clinicians, for example, were required to weigh up risks against benefits. Hauer *et al*, (2015) determined that trusting an individual was to ‘assign a responsibility or put something into someone’s care’ and entailed ‘dependence on something future’; a prediction. Allowing learners to practise independently too early, for example, could impact on standards (e.g. patient safety) with a resulting effect on supervisors and their employing organisations (e.g. liability for supervisors and costs for the healthcare organisation). Conversely, allowing learners independent action could positively affect their confidence, learning

curve and timely achievement of competence. There was, therefore, a balance to be struck between the ‘destructive friction’ of the former and the ‘constructive friction’ of the latter (Govaerts, 2017).

In this respect, trust decisions could also be regarded as ‘calculated risks’ on the premise that adverse events were manageable (Hauer *et al.*, 2015). For example, a supervisor might trust that a trainee was able to manage patients independently in an adjacent room most of the time because the supervisor could be there quickly enough to assist the trainee if a complicated issue arose. Studies have found that supervisors sometimes did not observe trainees directly and tended to explicitly or implicitly make estimations of the trainee’s adaptive competence to cope with unfamiliar situations. This suggested that trainees were sometimes judged by supervisors on activities that supervisors had never seen or that trainees may never have encountered (Hauer *et al*. op. cit.). Gingerich (2015), posited that trust judgements were likely to incorporate both deliberate and automatic social cognition processes (e.g. reflective and empathetic responses). Other studies found that experts in agriculture, personnel selection, healthcare, accounting and management had generally been unable to articulate the processes they used to make trust judgements or verbalise the thoughts behind their decisions (Shanteau, 1988). When asked to describe their decision strategies, experts would often refer to vague concepts. The nature of surgical judgement, was therefore an important research question for this study, asking how surgical supervisors made trust judgements enabling trainees to perform with increased levels of independent practice.

4.6 Literature on clinical trust

Traditionally, the concept of judgement was closely aligned with the conferral of trust, based on the notion that the professional was ‘infallible, altruistic and in possession of unique expertise’ (Lunt, 2008, p.76). This study, however, focused not on the trust conferred on the profession by the public but rather on how professionals, in supervisory roles, exercised trust on behalf of the public. In the training arena, the public’s trust in the clinician extended to the evaluation of how far trainees were deemed trustworthy to practise independently. As such, clinicians were not only trusted, they were also *trusted to trust*. Therefore, in the clinical context, the concepts around judgement, expertise and trust were interdependent; the notion of trust framed clinicians’ expert judgements leading to decisions that enabled trainees to practise. The underlying principle was that the greater the level of the supervisor’s expertise (built up over time from experience), the greater the likelihood of accurate trust judgements leading to successful patient outcomes. This interdependent relationship is summarised in figure 3.

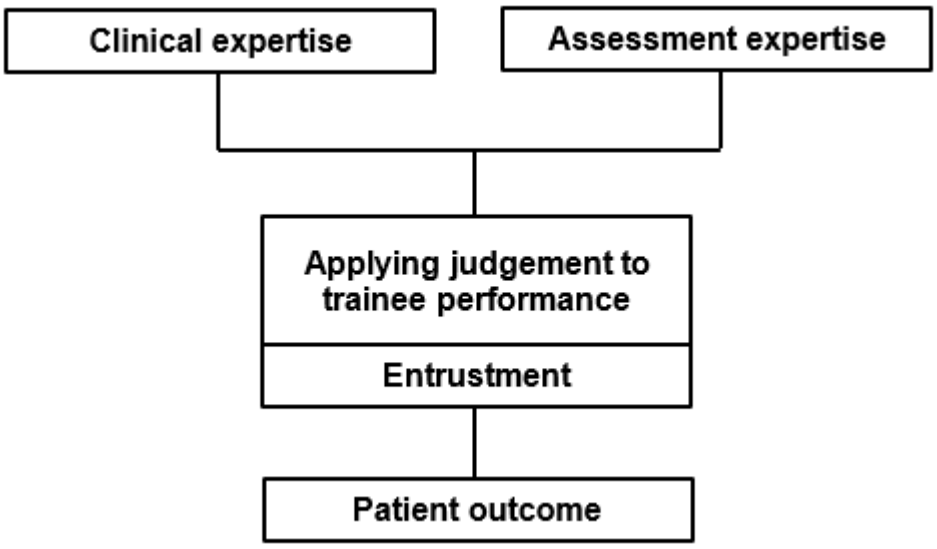

Figure 3: The use of expertise in the judgement of trainee performance

The use of trust was asserted by the proposed curriculum as a vehicle for delivering a more meaningful and intuitive taxonomy for practice than a mere list of competencies. Gingerich provided a general definition of trust as a social judgement which could be ‘inferred, felt, created, discovered, earned and lost, but not observed’ (2015, p.751). The proposed surgical curriculum advocated that trust could also be awarded (as a supervision level) as outlined in section 2.2. However, its impalpability created a problem concerning how it could be reliably evaluated. Trust judgements were associated with expectations of favourable outcomes (e.g. better trainee grading and patients’ well-ness) and emotions of hope and confidence (e.g. in successful training). In contrast, distrust was associated with fear, scepticism, cynicism, hesitance, increased wariness and vigilance (such as a supervisor’s unease and disposal to prompt a trainee or intervene). Trust could also be defined as ‘a willingness to be vulnerable to the actions of another party’ (Mayer, Davis & Schoolman, 1995, p. 712; Holzhausen *et al*., 2017) and as such was a risk-taking activity for supervisors as the ‘trustors’ (Damodaran, Shulruf & Jones, 2016; Gingerich, 2015; Govaerts, 2017). Supervisors, for example were accountable for decisions that impacted on patient safety. Supervisor judgements could typically arise from a sense of unease and take the form of increased vigilance of trainees even when there were no specific indications to justify such concern (Gingerich, 2015). In recognition, Gingerich proposed that assessment might require a shift away from directing supervisors to focus on purely cognitive activities such as ‘knowing’ or ‘recognising’ towards encouraging them to report feelings more congruent with trust, such as feeling calm while trainees carried out a procedure or alternatively being on edge.

When it came to decisions about trust, Choo *et al.* (2014), found that the development of trust was multifactorial and highly contextual. This notion corresponded with the view that a decision to trust was ‘always contextual and specific’ and that each scenario brought into consideration ‘variables beyond trainee competence’ (Damodaran, Shulruf & Jones, 2016 p. 338; Mayer, Davis & Schoorman, 1995). An influential study using entrustment of professional activities in the field of anaesthesia (Sterkenburg *et al*., 2010), explored the factors that influenced trust decision-making and concluded that four main categories of factors affected decisions about whether and when a trainee was ready to execute a critical activity independently. None of these four factor groups acted independently but represented discrete and measurable constructs:

- factors relating to the trainee
- factors relating to the supervisor
- factors relating to the task at hand
- contextual factors including socio-cultural aspects

As well as perceived ability, *factors in the trainee* included benevolence (extent to which they were believed to want to do good), integrity (such as adhering to a set of principles that the trustor found acceptable), recognition of their own limitations and self-efficacy. Decisions were also based on personal characteristics, such as conscientiousness, the perception of honesty, disposition, prior experience and expressed future field of specialty. Barriers to trust included failing to seek help promptly and overconfidence. *Factors in the supervisor* included their experience, expertise, approachability and attitude to assessment (although consideration of the key elements of self-confidence and

peer pressure were omitted). As trustor, the supervisor's propensity to trust was thought to be a stable characteristic akin to a personality trait (like neuroticism or agreeableness). The *factors relating to the task* included difficulty and complexity, the risk of complications, and the condition of the patient. The *relational and contextual factors* included the time of day, availability of other personnel, time, the quality of the relationship between the supervisor and trainee and team culture. While these findings were broadly supported by Bolger and Wright (1994), Mayer, Davis and Schoorman (1995), Dijksterhuis *et al*. (2009) and Choo *et al*. (2014), the study omitted the crucial element of patient factors (the influences brought about by patient behaviour). While it could be argued that the condition of the patient was included in task factors, 'patient factors' could have been considered as a separate construct.

Govaerts (2017) also emphasised that trust decisions were influenced by political, financial and cultural factors in the organisational/educational context. In medicine, research findings showed that cultural values of autonomy in practice and the aims of maximising efficient delivery of healthcare services conflicted with a culture that valued taking time to directly observe and document performance.

Trust was, therefore, likely to be temporal, person, task and context dependent. It was likely to be influenced by factors related to the trustor, their experiences and their working relationships. It was also likely to be influenced by political, financial and cultural factors. Multiple factors and contexts, therefore, had the potential to influence the quality of trust judgements in different ways at different times.

On the development of trust decisions, Ten Cate *et al*. (2016) distinguished between three types of trust: *presumptive* trust (based solely on prior credentials without observation), *initial* trust (based on first impressions) and *grounded* trust (based on intensive contact with the trainee and on systematic data collection). The last resembled an ipsative assessment view (Hughes, 2014). If the supervisor did not have much contact with the trainee, he or she might rely on credentials or first impressions (presumptive or initial trust).

Studies in America found that a determination of trust could be rapid (hours to days) and were influenced by early interactions (Hauer *et al*. 2015). The development of trust was often uninformed by prior knowledge of the individual resident (job title equivalent to a trainee in the UK). Attending physicians (title equivalent to surgical supervisor in the UK) instead relied on their understanding of general performance standards. In the absence of sufficient direct observation, abilities such as organisational skills, knowledge-sharing and a professional disposition were often seen as a proxy for trustworthiness. Attending physicians' observations of clinical care could be informal and unstructured, comparing residents with what they themselves would do or would have done at that stage of training. They were more inclined to trust residents who showed concern for particular problems, correctly performed technical tasks, asked the right questions, generated appropriate management plans, picked up on non-verbal cues, had enthusiasm, worked well within the team and demonstrated good reasoning (Hauer *et al*., 2015; Choo *et al,.* 2014). Choo *et al*. also found that many attending physicians checked their opinions by independently verified information to inform their perceptions of trainee honesty, attention to detail and ability to follow orders reliably. They would also inquire among their colleagues

and other senior residents to gain information about their trainees' abilities and integrity. Such an approach would also involve supervisor judgements about the trustworthiness of these sources of information.

Curriculum developers across medicine emphasised that the supervisor-trainee relationship was at the heart of the ability to trust (favouring Ten Cate's grounded trust), suggesting that the longer the contact with the trainee, the better the supervisor could estimate the trainee's capability. In addition, a higher degree of mutual trust was enhanced by shared expectations and ongoing communication (Hauer *et al*. 2013; Hauer *et. al*., 2015). The duration and intensity of the training relationship were, therefore, important (Ten Cate *et al*., 2016; Mayer, Davis & Schoorman, 1995). However, this model could be problematic if training relationships were of variable quality and when supervisors were not selectively recruited and trained. Ten Cate *et al*. (2016), suggested that there was often supervisor role ambiguity when responsibilities were described variously as coach, advocate or evaluator and that these differences had implications for a trusting relationship.

4.7 Literature conclusion

While studies have sought to understand the mechanisms that facilitated the development of expertise in different fields, they found that the highly complex and domain-specific nature of the task made it difficult to capture the deep application-orientated understandings of experts. While writers' categorisations of expertise, judgement and trust have helped to provide signposts to aid understanding, these typologies have varied significantly. An important division

in the literature stemmed from a disagreement about whether human experts were essential for higher order judgements or were flawed by the propensity to take shortcuts, in turn necessitating the use of more objective means. Proposed strategies to transcend the perceived cognitive limitations of human experts have varied between views that favoured the use of artificial systems and those that supported providing experts with criteria, decision aids and feedback.

Surgical curriculum developers were at the beginning of the journey to implement a system that aimed to better connect supervisor knowledge and experience with curriculum assessment. This study sought to provide some clues about how expert judgement currently worked. If supervisors were to use their expert judgements about trainees in an effective manner, the curriculum developer's fundamental task was to encourage them to engage in assessment activities that were likely to result in the achievement of learning outcomes by trainees. It was helpful to remember when carrying out this study that what the supervisor did was key to determining the design of the new assessment framework rather than the reverse. The related research questions asked how supervisors judged when trainees were ready for increased levels of independent practice, what types of judgement and cognitive processes they used and what factors/barriers influenced their judgements. In the interim, the literature summarised above suggested a useful combined definition of what supervisors did that when applied to surgical training might be helpful information for stakeholders. The new definition in the box below could also be relevant in other high-risk professions where expert knowledge was applied to training. This definition closely related to the definition of trust in section 2.1: The granting of permission to perform a function associated with patient safety.

New definition of supervisory judgement in surgical training

The exercise of supervisory judgement entails the undertaking of an important public task by those who have developed practical wisdom after much instruction, deliberate practice and experience. It involves a cognitive decision-making process that requires the active assigning of responsibility, based on a prediction, to a person with yet unproven skills.

The literature in this chapter established that expertise, judgement and trust were highly complex concepts that were also domain-specific. Any enquiry into the adoption of these concepts as a basis for formal assessment by supervisors needed to be designed around an understating of how supervisors tended to think. The literature also helped to define the critical qualities of knowledge, experience and deliberate practice that would be requisite for anyone undertaking this role but also revealed a dichotomy of belief about whether human expert judgement was preferable to adjudication through more objective, mechanical means.

The next chapter explores how the study was designed to create knowledge about the nature of supervisor judgement. It explains the values, responsibilities and assumptions made by the researcher about real world surgical practice and how these shaped the choice of methods and techniques that were considered optimal for answering the research questions.

Chapter 5: Research design and philosophical foundations

> ‘Just as the dancer relies on the spine for the power and coherence of the dance, so the qualitative researcher relies on the design of the study. Both are elastic. Like the dancer who finds her center from the base of the spine and the connection between the spine and the body, the qualitative researcher is centred by a series of design decisions. A dancer who is centred may tilt forward and backward and from side to side, yet always returns to the center, the core of the dancer’s strength’.
>
> (Janesick, 2018, p. 39)

Essential to the study design was its focus on a ‘real world’ practical problem and a qualitative and flexible approach. Flexibility was seen as a key component of qualitative research, combining as it did an evolving design, the presentation of multiple realities, the researcher as an instrument of data collection, and a focus on participants’ views (Creswell, cited in Robson, 2011). This chapter outlines the guiding paradigm for the study design. It sets out the role of the researcher and the choice of the ethical, theoretical, ontological, epistemological and methodological perspectives concerned.

Robson defined ‘real world’ research (2011) as referring to small-scale research projects that tended to be related to the evaluation of service or policy changes, carried out by individuals or small teams and where the need for answers fell within a short timescale. As the term suggested, the focus of real world research tended to be on practical problems, grounded in a specific context which had ‘direct relevance to people’s lives, to help find ways of dealing with the problem or of better understanding of the issue’ (p. 4). The notion of real world research had implications for how the study would afford deep insights into the community

of interest and had importance for the choice of methods, approach and language to be used as well as preferred ways of working. Also, as suggested by Hakim (cited in Robson, 2011, p. 71), it included elements of the researcher's own preferences and ideas as the 'architect' of the study as well as the 'stylistic preferences of those who pay for the work and have to live with the final result'.

An emphasis placed on the intention behind real world research by Robson was that it carried with it the 'suggestion of breaking out from the ivory tower' (p. 4). In this study, that involved affirming the researcher's intention of disconnecting herself from the area of curriculum design (the high ground suggested in chapter 1) and entering into people's natural settings. It involved using the verbal accounts of volunteer participants gathered in their natural settings and focusing on meanings in their context. Of the two principal research paradigms; quantitative (relying on numeric and statistical data) and qualitative (relying on narrative data), the study aligned in large part with the latter because qualitative methods were more responsive to natural contexts and could be used to explore processes that were embedded in practice.

The study favoured a social research perspective over the scientific and experimental tradition of the majority of surgical studies. It, therefore, limited the use of numerical data to the recognition of frequencies of statements in narratives while giving them less weight than the interpretation of what was important. It also accepted that the subjectivity and values of the researcher and participants were part of the study. Unlike fixed designs, appropriate to quantitative studies, a more flexible approach allowed the study to be carefully planned at an early stage of the process while being open to a certain amount of evolution as the research

proceeded. It also allowed theoretical ideas and concepts to emerge through an inductive logic. The two main risks involved with the design of real world research lay first in generalising too far from small-scale study findings and second in causing harm to people as a side effect of the study. It was important, therefore, that the study would seek to draw examples from the practice of participants rather than generalise findings to a broad context and would seek to recognise and limit the negative consequences on the people involved. However, consistent with all styles of research, the overall design aimed to ensure the study would be carried out in a systematic and principled fashion.

5.1 Researcher role and position

An important aspect of the research design was the researcher's 'situatedness' within the research context. Costley, Elliott and Gibbs (2010), determined that social situatedness arose from the interplay between the researcher (their unique perspective), the situation (the circumstances and researcher's position) and the context (the subject, background, problem, place and time). These elements affected the way a piece of research was undertaken and helped to categorise the researcher's status as either an insider (belonging) or an outsider (not belonging) in terms of those they were researching. Because of the unique perspective of the researcher, neither position could be truly impartial but there were advantages and disadvantages to each of these distinctions. The insider might tend to achieve better access to information and people and build a better rapport with those who saw them as having a shared citizenship. This had the potential to increase their perceived trustworthiness so that participants were willing to share information more easily. However, insiders might tend to find it

difficult to distance themselves from the expectations of those in their community, see emerging information from a fresh perspective and remove themselves from the study once it was completed. In contrast, an outsider might be viewed negatively as a stranger or someone with a conflicting purpose or connection who might, therefore, be trusted less or, alternatively, seen positively as taking a neutral view. However, according to Naples, insiderness and outsiderness were fluid social locations that were constantly shifting as part of an interactive process that 'negotiated embedded local processes' (1996, p. 84). As such, researchers were never fully inside or outside their communities of interest.

Because my study was placed on the 'high ground' of curriculum policy, it could be described as 'outsider' research as distinct from the more usual 'insider' research conducted by medical professionals working in the 'swamp' of hospital practice (Schön, 1990, p.3). This perspective may have helped me discharge some of the drawbacks of 'insider' research such as the risk that knowledge of the participants and their environment might have influenced the gathering of data. However, I could not claim to be an objective 'outsider' as I had a different type of 'insider' privilege, including access to surgical trainees and trainers, their data and direct involvement with related governance.

In contrast with the bi-polar positions this study might alternatively be described as the product of 'semi-insider' research which I defined as situated within one's own work practice and professionally allied to the practitioners of interest. As a semi-insider I occupied a tripartite position (i) as evaluator/researcher (ii) as a curriculum developer and (iii) as a learner. Inherent within each position was potential for ethically relevant role conflict as well as convergence. Figure 4

depicts these positions, linking each with the relevant stakeholders and role expectations that had to be satisfied.

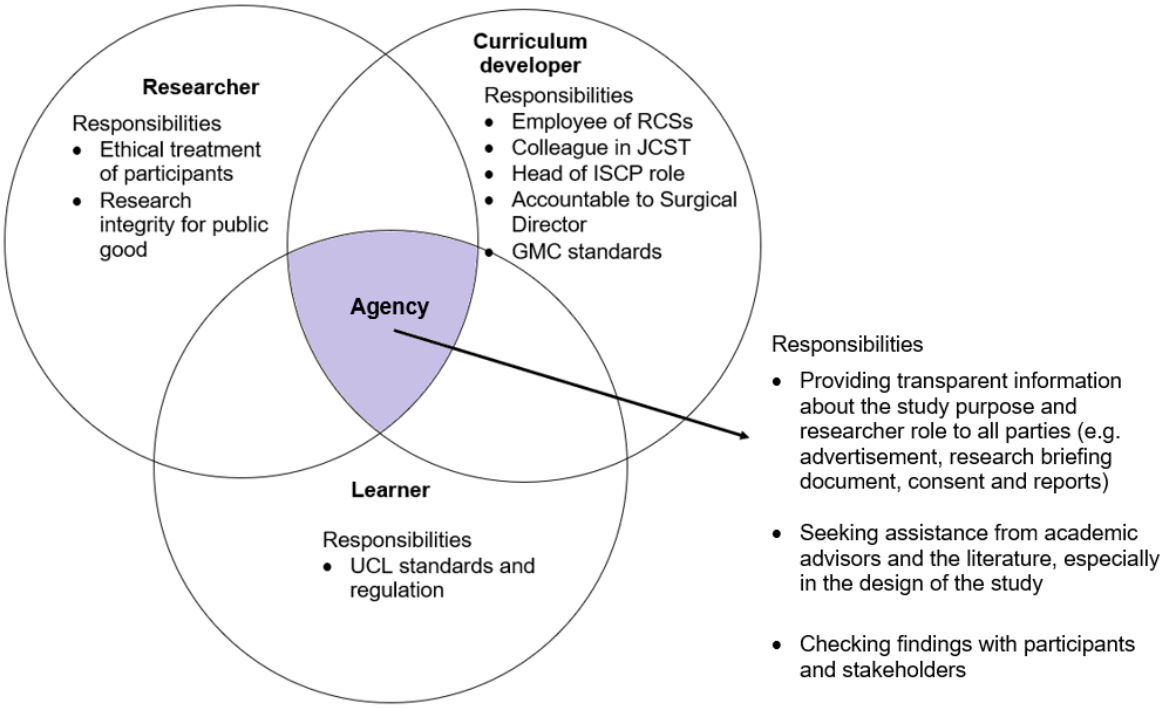

Figure 4: Researcher position and agency

I could act freely within the boundaries of the mauve area and this awareness allowed me to be honest with participants. For example, while conducting interviews, I could not deny my relationship with the surgical royal colleges and responsibilities to curriculum development, any subterfuge would have been likely to bring about unrealistic expectations, confusion and disappointment among different stakeholders at different points in the study. As a semi-insider I hoped to blend the insider advantages of easy access to an organisation and its people, as well as the understanding of the research setting and its wider context. This position was likely to be useful and relevant while embracing the opportunity for a fresh perspective that might arise from a ‘naïve’ questioning approach of an outsider. The ability to interpret participant views and behaviours would, as a

result, be a product of my research positionality, constructing one unique reality of the situation and, as such, the study would be an ingredient of the research itself. I hoped ‘insightful interpretations would emerge which would resonate with practitioners while acknowledging that there could be no universal “truth” claims’ (Wetherell, Taylor & Yates, 2001). My ‘semi-outsider’ perspective also gave primacy to pragmatism, an approach expected of me by those for whom I worked and described by Denscombe as taking the research problem as the starting point and ‘gauging the value of any particular approach in terms of how well its outcomes work in practice’ (2010, p.128). An implication of a pragmatist philosophy was that it allowed the choice of eclectic methods and aimed to address practical problems, providing those methods reflected the context within which they occurred.

5.2 Ethical perspective

Real world research involving human participants and the ethical responsibilities of my research and employment positions resulted in specific ethical considerations within the study. The study was guided by the Ethical Guidelines for Educational Research (BERA, 2011). My concern was to mitigate the potential for harm to participants, and through them, those for whom they were responsible. The UCL regulations were also clear on the need to ensure the accuracy of claims and citations, honesty and transparency, acceptance of responsibility of one’s own work and keeping up with related research. I viewed the researcher’s responsibilities as needing to help protect the credibility of the methodology in use and, as asserted as a key responsibility by Yin (2013), to

strive for the highest ethical standards. Ethical considerations were, therefore, important in ensuring roles, responsibilities and goals aligned.

Although the ethical considerations of the study were interwoven with the relevant sections of this study, social research tended to be value-laden (Robson, 2011), and it was, therefore, important to clarify the underlying key values and judgements that guided the study design. The four domains of principlism (autonomy, beneficence, nonmaleficence, and justice) served as a useful conceptual framework of headings to set out these values and also the expectations that the research sponsor, the Royal Colleges of Surgeons, had of its members, fellows and staff. Principlism appealed because it had a moral dimension to decision-making and was pluralistic in terms of compatibility with multiple ethical, theological and social approaches (Belmont Report, 1979).

Autonomy

My position as a semi-insider would have an impact on the researcher-participant relationship. As a non-practitioner who was not working in the NHS, I did not have dealings with the participants or colleagues they might know and no influence over their responsibilities. My employer, the Royal Surgical Colleges, also held no direct influence on their employment or appraisal. However, a risk concomitant with my responsibility for curriculum policy was that participants might have regarded my position as attempting to endorse the way they worked, a view which would run counter to my intention of placing expert supervisor judgement at the heart of the assessment process. With these risks in mind, participants in the study were contributors rather than subjects, autonomous agents who could opt in and out at any stage. I wanted to build trust through the process of individual

interviews and the provision of advance information about the proposed new assessment. My intention was to use participant perspectives to throw further light on what had been discovered through the current literature and help generate fresh theories to improve curriculum assessment in a way that would make it more pragmatic. Participant views were, therefore, seen as legitimate as guiding the study outcomes. Balanced against this, my design and leadership of the research on the behalf of all the stakeholders connected with my role (as outlined in figure 4) meant that my research decisions had the potential to directly impact on training policy although participant contributions could impact on the study results.

An essential aspect of the study was its independent questioning and impartial interpretation. It was important that I took a neutral position in relation to the proposed new framework, neither promoting nor opposing it. The framework was couched in terms of a GMC initiative that was open to modification or challenge rather than being pre-determined. As a result of this undertaking, participant feedback had to be taken into account for subsequent modification of the assessment content, policy, procedures and faculty training. Related to the participants' ability to feed back, was the notion of validity in constructionist social research. The study would provide opportunities to present the researcher's findings back to the participants 'to check convergence with the experience of the researched' (Edwards & Furlong 1985, p. 33). A sense check of this sort would ensure that any claims and conclusions were appropriately placed and where necessary, speculative.

It was important that participants were not misled. To ensure they understood the implications of joining the study (why their participation was being sought, how their data and accounts would be used, how data would be stored, for how long, how and to whom results would be reported and how the results would be used), they were provided with briefing information. Information about the study appeared in advertisements, consent sheets and the simulation exercise. Each participant would also have this information explained to them in person on commencement of interviews. It would include information about the right to withdraw from the study at any time without the need to provide a reason. The study took place in the highly complex arena of surgical training in the NHS. These situations tended to be inevitably political such that acceptance by stakeholders was by no means a given. Participants would be asked to sign a voluntary consent form to confirm this understanding before any information was taken, including the right to withdraw at any point. On completion or withdrawal from the study, the participants would be able to provide feedback, suggestions or concerns about their interview experiences. They were also informed that they could have transcripts of their interviews which would be anonymised.

Beneficence

A predominant goal of the study was to stimulate in supervisors the values and behaviours espoused by the curriculum, aiming to simultaneously enhance the supervisory role for them and improve the curriculum. I hoped to arrive at recommendations for improvements that would enhance supervisor performance and add to the body of knowledge about assessment. I, therefore, took a long term and non-minimalist view of my responsibilities to the conduct of the study, its participants and the wider surgical community, aiming to create foundations

for further developmental work. The needs of the participants took primacy over personal and curriculum goals, and this stance was consistent with my organisation's position as a member body and a responsible entity. As a result, I was prepared to challenge the assumptions behind the curriculum and demonstrate willingness to confront the perceived weaknesses of the proposed new assessment framework. It was important that the participants' contribution would benefit them directly. A cost to the supervisors for joining the study would be the time taken for the interviews and simulation exercise and their divulging of personal training information. To mitigate this, I needed to conduct the evaluation with a minimum of invasiveness. The anticipated incentives for supervisors would be the opportunity to influence future assessment, to obtain advance information about forthcoming regulatory change and evidence of their involvement in educational activity. It was likely that the study would have been of most interest to surgical supervisors from the specialties closest to the planned live implementation date (2018 at the time of starting the study).

Nonmaleficence

The study had minimal potential to adversely impact on the participants' practice or local processes because of their high-level positions as consultants responsible for their own surgical units. I endeavoured to make the experience enjoyable, easy and educational and to acknowledge their contributions professionally e.g. by providing a certificate of participation. All data were securely stored and processed in compliance with the Data Protection Act (Gov.uk, 2018) and the updated General Data Protection Regulations (Ico.org.uk, 2018). To respect anonymity and privacy, participants and those to whom they referred were anonymised. Only generic descriptors were used to identify their

characteristics e.g. broad regions rather than hospitals. Supervisors were asked to talk about the trainees they worked with in an anonymised way e.g. ‘the ST5 trainee’, ‘the junior trainee’ etc. Transcripts from each participant would not be shared beyond the sole researcher. Audio and transcript files of participant interviews would be deleted on completion of the study.

Justice

I sought to use rigorous research methodology, validated by my UCL academic adviser and to conduct it fairly and transparently e.g. through open recruitment, providing participants with good information and appropriately acknowledging the participants’ contribution throughout e.g. in presentations and publications. Reports would not provide any pre-conceived position or sway towards supporting evidence but rather be open to reporting contrary evidence.

5.3 Theoretical perspective

According to Crossman (2018), a theoretical perspective is a set of assumptions about reality that inform the questions we ask and the kinds of answers we arrive at as a result. In this sense, theoretical perspective could be understood as a lens through which to look, serving to focus or distort what was seen. This study relied on several assumptions about how judgement occurred in surgical practice. At its heart was the notion that supervisors’ judgements were historically and culturally located and, therefore, subject to variable behaviours, attitudes, experiences and interpretations (O’Gorman & MacIntosh, 2015). Supervisors, therefore, developed their practice from their unique background, personality and experience, interactions with trainees and other trainers, circumstances of their

practice and the nature of the task. These notions would have a bearing on the creation of knowledge e.g. about how to judge trainee performance, through the transformation of experience.

The aim of the study was to develop a theoretical understanding of the phenomenon of judgement and not be constrained by existing theory. However, Kolb's spiral of experiential learning was a useful initial theory or conceptual framework for assumptions made about supervisor judgement as a 'process whereby knowledge is created through the transformation of experience' (1984, p.38).

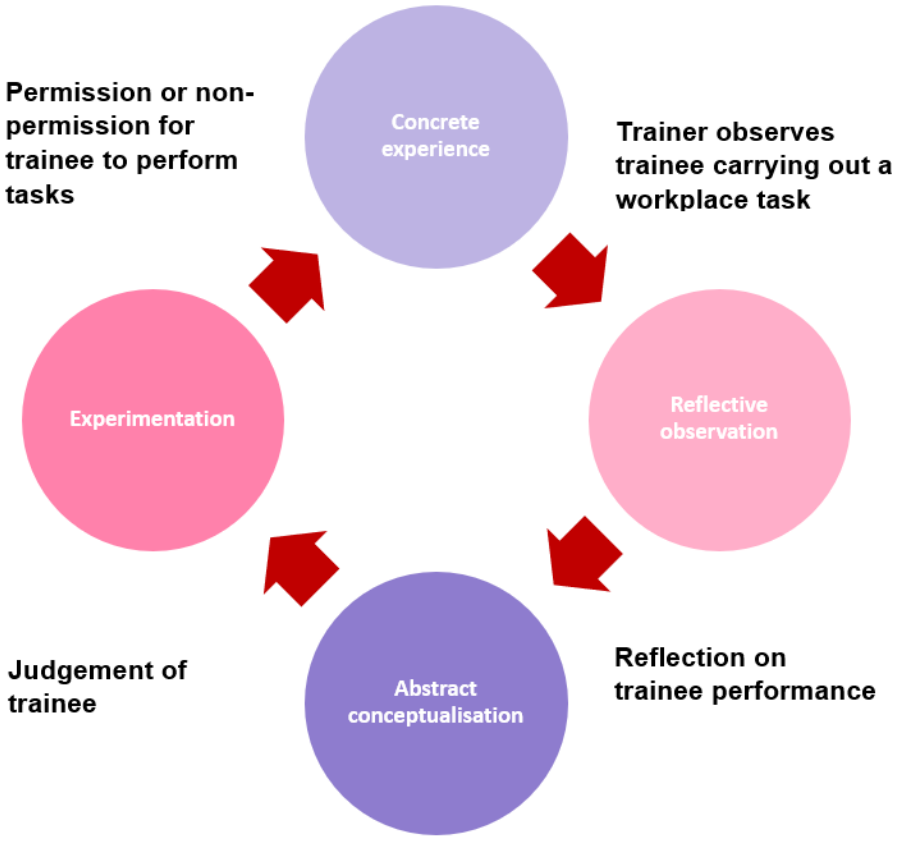

Figure 5: Kolb learning cycle theory related to supervisor judgement

Drawing on Kolb's theory placed emphasis on the supervisor's internal processes, supposing that knowledge, skills and understanding were acquired

and retained through cognitive, emotional, and environmental states as well as prior experience. When related to supervisory judgement, the cycle of learning was initiated by training situations with trainees that provided ‘concrete experience’ which resulted in the ‘abstract conceptualisation’ about the performance of trainees, leading to ‘active experimentation’ in their judgements, in turn creating new experiences. Kolb’s cycle, illustrated in figure 5 on the previous page, allowed that each supervisor experienced their place and time in the world in a different way. When interacting with other people they were constantly making subconscious comparisons to ascertain their position within that interaction and this mechanism was on-going throughout training. The study would have the potential to explore how far these established practices were subject or open to change.

5.4 Ontology

Denscombe defined ontology as referring to beliefs about the nature of social reality (2010, p.118). According to O’Gorman and MacIntosh (2015), there were two basic ontological positions; subjective and objective. This study saw judgement as subjective, conceptualising it as being constructed through the individual perceptions of different supervisors and reinforced by their interactions with other people. However, the notion of objectivity was not rejected outright. The study recognised the importance of human creation of meaning. Judgement was seen a process of meaning-making wherein surgical supervisors worked within a system and had an important and influential role to play. Surgical practice was viewed as having evolved as a unique culture comprising multiple realities and knowledge was constructed collaboratively within the social structure of

surgical training. Kolb's stance suggested that learning was an individually tailored process of construction. His theory extended to the belief that learners had preferences for different learning styles with strengths and weaknesses in different points on the learning cycle influenced by their own basic cognitive structure and experiences. The educational implications were that curriculum and assessment had to be designed to accommodate different judgement patterns through each stage of the learning cycle. The epistemological implications for the study were that each stage of the learning cycle had to be explored from different perspectives if curricula were to be aligned and improved. This position reflected an intention away from behaviourist and positivist views of learning which supported, for example, either the top-down introduction of curriculum or a system of grades and targets in education.

5.5 Epistemology

Epistemology refers to the ways that humans create knowledge about the social world and concerns the logic behind the ability to acquire knowledge of social reality (Denscombe, 2010). To gain knowledge about the judgement of supervisors, the locus of the study was on supervisor accounts of their practice. There was potential for gaining insight, through an interpretivist lens, into how supervisors saw the world (Burr, 2003) e.g. what they thought was important and not important, and to gain an understanding of how the environment of surgical training shaped their judgement behaviours. An interpretivist philosophy would seek to gain a contextual understanding of what was happening in terms of meaningful categories of human experience. It would aim to take into account the multiple realities revealed by the perspectives of different participants and make

an interpretation of the collected data. The use of both subjective accounts and an interpretivist approach aligned with a qualitative paradigm, consistent with a concern for the way in which people (researcher and participants) shaped their world and which Robson defined as focusing on ‘human beings in social situations’ (2011, p.17). This study, therefore, placed in high regard the inner world and experience of supervisors as a means of discovering how they used judgement in assessment and how assessment could be brought more into line with their judgements.

5.6 Methodological perspective

This section explains the processes and methods relating to the chosen methodology (case study), through which the research questions were addressed. Because judgement was seen as subjective and the result of the transformation of individual experiences, personal experiences were sought. Denscombe (2010) argued that methodology followed on from research questions and Yin (2013) asserted that ‘how’ and ‘why’ questions favoured a case study approach. According to Baxter and Jack (2008), qualitative case study methodology offered a valuable approach for the study of complex phenomena within their contexts such as the health sciences. It could be used to develop theory, evaluate programmes and develop interventions. This study sought to show how supervisors judged and how their current judgement patterns compared with a new assessment approach. Following the lead set by Yin (2013) which allowed that case study methodology could be exploratory (as well as descriptive and explanatory), the study explored the accounts and ideas of supervisors to achieve an understanding of their judgement processes. The

knowledge gained, therefore, derived from listening to what they said about how they interpreted their realities. According to Yin (op. cit.), a case study design could be considered when it was not possible to manipulate the behaviour of those involved in the study, when there was a need to cover contextual conditions because you believed they were relevant to the phenomenon under study and when the boundaries were not clear between the phenomenon and context. To give an authentic picture of supervisor judgement, the context of NHS trust hospital settings had to be considered as it was in these settings that the judgement decisions were developed and employed. Yin's approach to case study methodology aligned with a constructivist (similar to interpretivist) philosophy, built on the premise of a social construction of reality. The approach gave prominence to close collaboration between the researcher and participants, enabling participants to relate their views of reality and allowing the researcher to better understand participants' actions.

The unit of analysis

In case studies, a 'case' as the 'unit of analysis' (Baxter & Jack, 2008) might be an individual, a situation, a group or an organisation (Robson, 2011). Because the study was interested in focusing specifically on the experiences of supervisors, each supervisor was an individual case set within the real world context of surgical training and assessment. Baxter and Jack also placed emphasis on setting boundaries about the scope of case studies in terms of place, time and activity to avoid the problem of answering questions that were too broad and, therefore, the cases were set in the situational context of summative assessment at a real trainee progression point. Yin (2011) distinguished between single and multiple case study design decisions which had

to be made prior to data collection. The choice was related to the theoretical propositions of interest underpinned by the research questions. The rationale for each supervisor forming a single case within their work setting was that each was critical to the theory of judgement being explored because each supervisor could represent an extreme, unusual or a revelatory pattern of judgement, offering a distinct opportunity worth documenting. Conversely as single cases each supervisor could also provide evidence of commonality. Baxter and Jack, drawing on Stake (1995), used the term ‘intrinsic’ when there was genuine interest in the understanding of the case in all its particularity or ordinariness rather than because it represented other cases, a particular trait or problem.

An instrumental case, in contrast, provided insight into an issue or helped to refine a theory and, therefore, a secondary, instrumental aim of the study was to compare judgement practice with the proposed new assessment. The study design involved introducing a simulation of a new assessment framework to participating supervisors and asking them to use it as a lens through which to reconsider the performance of the trainees working under them. The study explored their thinking about their performance decisions, probing how far their perspectives on their trainees might have altered through the new framework (e.g. whether a trainee judged as highly performing through the current competency assessment regime was seen as embodying the same or different qualities identified through the new framework) and whether different levels of supervision could be employed to stratify different levels of performance. In this way, the study simultaneously considered how far each supervisor was able to assimilate the framework.

Generalisation

According to Yin (2013), there was a tension between the unique, contextually specific nature of a case and the need to make sense across a number of cases. Through case comparison, it was possible to explore convergent themes that might emerge and, where possible, draw 'analytic generalisations' which differed from statistical generalisation drawn from empirical studies. When lessons from different case studies were compared, a common explanation could emerge to categorise problems. Through analytic generalisation, there would be no inference made about the study population on the basis of the data collected from the small number of participants in the study i.e. generalising findings beyond the sample of participants and applying them to the larger population of all consultant surgical supervisors. To do this would be a 'fatal flaw' in the methodology as described by Yin (p. 40) because the cases were not 'sampling units' and too small to represent the larger population. Instead, the participants' responses would be considered to be an 'opportunity to shed empirical light' on the research problem. For example, findings could be used as a working hypothesis either to be applied to reinterpreting the results of existing studies or new research focusing on nuanced situations. Therefore, the aim was still to generalise to other concrete situations as well as to a variety of like-situations represented by the original cases.

According to Yin, study propositions could help to form the groundwork for generalisation. The propositions in this study were set out through the research questions and could be generalised as shown in table 5. Alternatively, with a flexible design, new generalisations could emerge from the study findings. The

main aim of generalisation would be that it would be set at a conceptual level higher than that of the specific cases.

Generalised research questions		**Propositions**
How do supervisors identify when learners are ready for increased levels of responsibility?	• What factors influence supervisors' judgements?	• Objective, structured competency-based assessment regimes
	• What cognitive processes do supervisors use to judge learners?	• Kolb cycle of experiential learning
	• What patterns of expert judgement occur in practice?	

Table 5: Adapted research questions to aid analytic generalisation

The case study approach had the potential to support the study in three ways:

- To help gain understanding about different supervisor processes for making judgements and the factors that influenced them;
- To support the introduction of a new assessment by providing information about how supervisors understood and used the simulated version;
- To provide broader analytic generalisations about the nature of supervisor judgements.

Summary of methodology

In summary, the case study design offered an opportunity to gain insight into the practice of a small number of surgical consultant supervisors working in complex situations. It involved exploring the strategies they used to make summative judgements about trainees and the degree of autonomy they allowed them, using as examples their current trainees and the context within which they worked. To

counter the concern that case studies were not always seen as systematic or faithful to careful design (Yin, 2013; Robson, 2011), it was important that the study had rigour in design, data collection, analysis, interpretation and reporting. For example, case study procedures had to allow for checking of reliability and replication by other researchers through documented case study protocols and were a way of increasing the reliability of case study research. The table below summarises the research design construction.

Qualitative research paradigm	
Ontology	Subjective / Constructivist
Epistemology	Interpretivist / Accounts / Natural settings
Data gathering	
Methodology	Case study / Exploratory
Techniques	Interviews / Simulation

Table 6: Summary of research design

The perspective adopted in this chapter was associated with real world social research, focused on practical problems that were grounded in specific contexts and related to improving the understanding of an issue. It was argued that a qualitative approach was needed to explore the way in which supervisors applied their judgement. This design had implications for how the study would collect and analyse data in the next chapter. Chapter 6 extends this set of beliefs toward an inductive approach to data collection and an interpretive approach to its analysis with the aim of identifying patterns of behaviour that had developed through experience.

Chapter 6: Data collection and analysis

> "To find signals in data, we must learn to reduce the noise – not just the noise that resides in the data, but also the noise that resides in us."
>
> Stephen Few (2015)

6.1 Data collection

When considering how to undertake data collection to answer the research questions, two alternatives presented themselves; either to observe practitioners at work or ask them about their work. The former would have taken a great deal of time because of the opportunistic nature of supervision and assessment in surgical practice. Observation may also have resulted in misinterpretation of the words and actions between trainee and supervisor because of the unique context within which each training situation occurred. For pragmatic reasons, therefore, the mechanism for data collection was through supervisor accounts of their work, facilitated by semi-structured questions in a private space in the participants' work settings.

Semi-structured interviews had several advantages, allowing there to be a list of key topics or questions that needed to be covered but at the same time allowing exploration guided by the interviewee's responses in a conversational mode (Robson, 2011). According to Yin (2013), interviews were one of the most important sources of case study evidence provided they resembled guided conversations rather than structured questions. The study's interview topics covered supervisor values and how they adopted them, what helped and

hindered their supervisory role, how they made and acted on judgements, how they perceived the capability of trainees and managed those who were above and below the norm (see appendix 3).

Participant selection

Through the selection of six supervisors, I intended to provide a suitable variety of compelling examples of similar or contrasting patterns of judgement behaviours and to evaluate how far they aligned with the propositions offered by the new assessment framework (table 5). The aim was to explore different supervisor strategies around judgement about the performance and readiness of trainees to perform different aspects of practice and the implications of introducing a new approach.

The vehicle for recruitment was an open advertisement on the curriculum website (www.iscp.ac.uk), an approach that worked well for my Institution-Focused study (IFS). This medium provided a fair means of reaching all supervisors in higher surgical training with the risk that it would tend to be noticed by those who were most engaged with training, and, therefore, visiting the site most often. The primary condition for selection was to find individuals who represented different training experiences and who might, therefore, have developed different judgement patterns. This would include people from different backgrounds such as those who represented the protected characteristics set out in the Equality Act (Gov.uk, 2010). In surgery, the two most common minority groups included women (gender category) and surgeons from non-UK cultures (race category). The secondary condition comprised three factors, to find i) supervisors in different specialties, ii) in different regions and iii) with varying experience (including

responsibility for at least two trainees). Supervisors were accepted in order of application providing they conformed to the above criteria. Out of sixteen interested supervisors, the six supervisors chosen ensured an element of randomness and a mixture of regions, specialties, backgrounds and gender (see appendix 2). The selected participants were sent a detailed briefing sheet about the study background and interview structure as outlined below.

Interviews

Using Yin's typology, interviews were used as 'prolonged case studies' (2013, p. 110), because they would take 2-3 hours in a single sitting or over the period of one day and be in three parts. It necessitated a large one-off commitment from participants, an unusual research mechanism which might have appeared to make unreasonable demands on these busy surgeons. However, interviewing on the same occasion reduced the possibility that other factors would distort supervisor views between interviews and added integrity to the gathering of multiple sources of data within the same time point. There were also advantages for the participants in terms of their time and energy. They tended to utilise their programmed time allowance for educational activity to work with me. Before the interviews began, each participant was given an overview of the study and written consent was taken (approximately 20 minutes).

PART 1: The first interview asked participants to describe how they made judgements about trainee performance and progression, exploring their experiences and feelings about what factors helped and hindered their judgements. They were invited to share stories and contextual examples about current practice (anonymising the names of their trainees and colleagues). They

were asked to describe what led to good and poor judgements, decisions they made and their reasoning for them. The goal would be rich contextual knowledge within each case study that could help the readers of the study visualise the situations in which different progression strategies were used.

PART 2: The first interview was followed by a practical exercise in which participants considered their trainees' progress using three key CiPs (covering the main surgical environments of ward, clinic and theatre, see Appendices 4-6). At the time of the interviews, the participants were unlikely to have advanced information about the new CiPs framework and, therefore, each was given a detailed briefing on how the new process would work, enabling them to ask questions (approximately 20 minutes). The participants were then left on their own to read through the CiPs before conducting the assessment with a focus on their own trainees. They could take as long as they needed. The practical exercise was considered to be a mock summative assessment faithful to the manner in which supervisors might use the new framework if assessing trainees at the end of a placement.

PART 3: The simulated assessment was followed by a second interview inviting participants to talk through their thoughts about their trainees in light of the three CiPs as a different lens on performance. Questions explored differences in the way they saw the progress of each trainee and the way they made their supervision level choices as well as their feelings about the new assessment, the problems and solutions they thought might transpire as a result of implementation and what guidance and support they thought they and other supervisors would need for successful implementation.

Interview considerations

The interviews were individual, face-to-face and semi-structured. These options allowed a conversation to be had, encouraged honesty and a logical questioning sequence. Rather than allowing unstructured freedom to talk on a broad topic, this style allowed an interview to stay focused on its purpose and time allowance with some flexibility for expression. Additional, unplanned questions could follow up or probe participant statements and could lead to questions being changed for subsequent interviews (for example, exploring areas in which trainees tended not to be observed). Being with the participants allowed reading of their non-verbal language while endeavouring to convey reassurance about my intentions through my body language. A balance was cast, therefore between a free-flowing conversation which might have led to better insights and the pragmatic following of a purposeful course for consistency. I hoped to make participants feel comfortable and valued by travelling to their workplaces, which were all UK-based, and this took time to set up and follow through.

In order to receive information openly, I had to be a good listener. I also needed to keep control of time while addressing the research questions. This involved tactfully directing the participants who were in senior positions and generally accustomed to being in control and who could use the opportunity to *tell* me what curriculum policy should be. As the interviewer, I needed to be aware of any cognitive bias, for example I had to avoid allowing someone I liked to talk more (halo effect) or someone who was controlling to talk more about what they felt they wanted me to know. The tension between these priorities had to be carefully balanced.

Timing

Interviews were held around a point in training when supervisors would normally be undertaking a summative assessment of their trainees. They were, therefore, well placed to describe their current trainees and ways of making judgements as well as providing examples to illustrate their approach, how they had reached conclusions and how they managed trainees at different progression points. The timing allowed for a valuable conversation but also had to fit around each participant's working timetable. Each participant's views would be critical to the success of the case study by providing individual insights. Therefore, no one participant would be more critical to the study than another.

Questions

Questions posed in interviews were important because various writers had shown that experts in different fields had tended to find it difficult to articulate their expertise (Bolger & Wright, 1994; Coles, 2002; Dreyfus & Dreyfus, 2005). Yin asserted that interviews pursued a line of questioning but also needed to be fluid rather than rigid (2011). The interview questions were designed to be constructive, asking not only about current behaviours, thoughts, practices and needs but also about ideas for change (see appendix 3). The questions were mainly open in nature, allowing in-depth responses e.g. *'Talk me through the process of how you do it'*. I used encouraging sounds as an indicator of being non-judgemental, allowing them to re-live training situations without fear that they would be judged on whether they were following the curriculum's espoused protocols for assessment. I used periods of silence to allow them time to think. Probing allowed for clarification e.g. *What do you mean by …?*' In addition, I had to be able to make keen observations while sensing what was going on in more

than just the aural modality and, therefore, questions were couched not only by tone of voice but also by the words used, altering the verbal form from the pure written form where appropriate. For example, sometimes a ‘how’ question was used to address a ‘why’ question to appear non-threatening such that the more direct ‘*why did you do that*?’ was replaced with the softer ‘*how did you come to do it in that way*?’. The initial questions were broad e.g. asking what the interviewee enjoyed about being a supervisor. To place the talk about judgement in terms of how it fitted with practice, an early question asked them to describe how they saw the culture of training and what helped and hindered their role. Talking about values allowed me to probe for specific examples of aspects of assessment that worked and did not work with the culture of training. Difficulties and barriers encountered e.g. the ability of supervisors to challenge the service environment, were of interest for my thinking about what shaped judgements. Because the new assessment purported to assess progression, I asked how progression was recognised and judged and whether and how trainees were allowed to progress at their own rates. In the second interview I asked the participants which elements of the new framework were feasible and what might help other supervisors adopt the method. I explored with trainers the demands of training, the time they spent on assessment, their comfort level with verbal and written elements and when and where feedback should be given. Following the guiding principles set out by Robson (2011), the questions were pre-tested with a non-participating supervisor and went through many iterations.

Interviews as a research method fitted with the qualitative design as they typically helped to reveal what people did and thought in some depth and allowed the participants some flexibility in the way in which they chose to respond to

questions. Balanced against this was the risk that participants might edit their responses to make them more acceptable. However, I viewed interviews to be more likely to engender rich perspectives than more remote methods such as a survey. They were my primary source for guiding further literature searches and generating theory.

6.2 Data Analysis

An exploratory approach to the data and the interpretation of what the data communicated as important were seen as key to the flexible design. The data analysis, therefore, involved what Yin (2013) described as a 'ground up' approach in which the data were closely examined rather than the preoccupation with predetermined theoretical propositions. This approach allowed key concepts to emerge in order that insights could initiate an analytic path and possibly suggest additional relationships. A systematic approach helped to assimilate the very large amount of data and reduce bias and ensure rigour. There were three specific stages to the study, allowing the analysis to switch between *closeness* to and *distance* from the data. The overall aim was to take an interpretivist approach, asserted by Yin (op. cit.), as taking into account the multiple realities revealed by the perspectives of different individuals, the context, the contextual understanding and interpretation of the collected data and the nature of the researcher's involvement so that the focus was on understanding rather than measuring.

Stage 1: Broad picture

Data from each interview for each participant were carefully transcribed in Microsoft Office Word documents. The broad view began with a simple read-through of all the interview texts without coding in order to obtain an overall impression of the richness of the data (i.e. the degree to which it gave an account of supervisor behaviours), to detect any immediate concepts and avoid the trap of getting specific too early. This stage helped the understanding of the context from which the interviewees perceived the training world and aimed to infer the meanings intended by the interviewee not the researcher. The goal for this stage was a general understanding of the scope of the data.

Stage 2: Formal coding

Because of the small size of the group, analysis of the data did not require specific coding software. The unit of analysis for coding was complete sentences which also provided the context for the use of particular words. Coding was data-driven rather than referenced to a pre-existing coding frame as it was important to the study that existing theories did not confine the thinking. Coding involved detailed, slow and reflective exploration of each text, asking what kinds of ideas were recognisable. The aim of creating codes (or labels for segments of data) and their descriptions (precise meaning of the code) was to 'break open' the text (Bazeley & Jackson, 2013, pp.72, 74). Some texts were 'sliced' so that multiple codes were present across a text e.g. "Can they [trainees] make a decision?" was relevant to three codes; '*Trainee qualities*', '*Clinical decisions'* and '*How do you know to trust?*'. This example was also relevant to a parent code e.g. '*Trainee qualities'* and then to a finer-grained child code of '*Decision-making'* in order to explore ideas rather than to take coding to its ultimate conclusion. Each code held one

concept (parent code) or a part of one (child) as directed by the data. Coding continued throughout each transcript and involved an iterative process of reading through the texts carefully and systematically line by line, highlighting sentences according to ideas. Techniques that helped to identify ideas and themes included:

- Repetition – e.g. participants used the words ‘tick-box’ to indicate current assessment practice.
- Indigenous categories / linguistic connectors - e.g. participants used the words ‘call me’ to describe a type of trainee decision and a type of supervision.
- Similarities and differences between participants - e.g. all participants valued sound management plans.

Through these ideas, I explored possible answers to the research questions that sought to identify what was done, what was valued and not valued, difficulties and solutions, differences and similarities. The analysis was, therefore, based on my understanding and interpretation of the meanings conveyed in the transcribed interview data.

Initially there was a rapid emergence of codes from the texts. When new codes were created by new sentences, previously coded texts were re-analysed to determine the presence of those codes too (constant comparative method). Returning to texts more than once was, therefore, necessary. The number of new codes decreased markedly after the first participant. The full transcripts were coded in this way. The result was a long list of codes across the dataset (appendix 8)

Stage 3: Theming

The development of themes involved the sorting and combining of different codes into categories and categories into themes with the aim of conceptualising supervisor judgement practice. According to Braun and Clarke (2006), theming at a latent level allowed exploration of the 'underlying ideas, assumptions and conceptualizations – and ideologies – that are theorized as shaping or informing the semantic content of the data' (p. 84), going beyond describing what was written to interpreting the features that gave accounts their particular form and meaning. Themes were, therefore, seen as categories conceived by the researcher that together formed the theory and outcome of the study. A large number of references did not necessarily give importance to a theme. Areas were considered as themes if they were thought to capture meaningful categories of human experience. There were three levels to the theming. In level 1 the codes were grouped according to subject matter (e.g. *Performativity* included codes where there were constraints resulting from a focus on targets and competencies). In level 2 the codes were clustered for similarity of content into broader categories (appendix 8). In level 3, the categories were further condensed according to judgement patterns that emerged (table 7).

Theme	Sub-themes	Key parent codes (level 1 & 2)
Judgement in Action	• Competence • Decision-making • Administration • Relationships	Nature of supervision
		How you know to trust
		Reasons for first time decision
		When you know to trust
		What might not be observed
		Standard
		Sources of evidence

		Trainee qualities (positive)	
		Nature of clinical decisions	Expert
			Environments

Table 7: Themes and codes

The theming process led to the identification of an umbrella theme under which four sub-themes were considered indicative of supervisor judgement of trainees and helpful in informing the practical development of curriculum assessment. These themes are developed further in the analysis in chapter 8.

As covered in this chapter, ideas from the participant accounts gave rise to patterns (or themes) within the data. The next chapter sets out the general findings from the analysis process, describing the factors that affected supervisors and the determinants of their trust judgements.

Chapter 7: Findings

> "Signals always point to something. In this sense, a signal is not a thing but a relationship. Data becomes useful knowledge of something that matters when it builds a bridge between a question and an answer. This connection is the signal."
>
> Stephen Few (2015)

A recognition of the factors that influenced supervisor judgement was essential to the study. This section outlines the influences that helped and hindered their judgements. It also covers the most important sub-theme from the theming process above – decision-making - describing the nature and quality of surgical decision-making and the notion of trust as the deciding factor in whether trainees were able to perform. It describes the determinants of supervisor trust, how trust was granted and what made trusting more likely.

7.1 Influences on judgement

The coding and theming process involved elements of judgement 'pattern matching' in a type of cross-case synthesis, exploring commonalities and differences in the strategies and thought-processes used by the participants to provide a better understanding of how participants tended to approach assessment judgements. The aim was to draw lessons that had implications for the development of curriculum assessment in surgery and beyond to form analytic generalisations where possible (Yin, 2013). From the coding process, it quickly became apparent that all the participants' accounts showed a very similar approach to judgement, an unanticipated finding in respect to the propositions

set out in table 5 (page 106). There was surprisingly little variation in how the participants judged trainees and precluded the aim of presenting separate cases of ways of judging. Instead, a more universal picture of judgement began to form.

Factors that influenced supervisor judgements

In relation to research question 1a), regarding influences that shaped supervisors' judgements, the study participants revealed that their training practice was shaped by cultural factors pertinent to the environment in which training took place. They frequently referenced three types of barrier which presented obstacles to a model learning environment but also shaped their judgements, these influences fell into three related categories; service, performativity and time. They also spoke about positive influences which were categorised in this study as *sources of evidence* (appendix 8).

The influences upon participants were considered important because of their potential to shape the process of judging and the judgements themselves. Kelman's theoretical framework (1958) had been highly influential in terms of promoting the central theme of social influence theory and still had currency. His theoretical framework proposed that attitudes corresponded to the way that individuals accepted influence (or conformed) which might occur at different levels. Kelman theorised that the three different processes of influence were *compliance*, *identification* and *internalisation*.

- Compliance – when an individual accepted influence because of social rewards and avoidance of specific punishments or disapproval. This influence

was more likely when there were controls in the environment and individuals were more likely to conform when they felt under surveillance.

- Identification – when an individual accepted influence to establish or maintain a satisfying self-defining relationship. This influence was more likely when the individual found the identity attractive and they conformed when they felt conditions were salient.
- Internalisation – when an individual accepted influence because it was congruent with his/her value system. This influence was more likely when the individual found credibility in a situation and they conformed when they felt it was relevant.

These levels of influence were useful in categorising some of the areas that participants felt played an important role in their judgements.

Level 1: Compliance influences

As noted in the study by Sterkenburg *et al.* (2010), contextual factors including socio-cultural aspects were important influences. Govaerts (2017) also set out that trust decisions were influenced by political, financial and cultural factors in the organisational context. Compliance influences included service, performativity and time outlined below.

SERVICE INFLUENCE: Participants faced constant tensions between their obligations for providing a service to patients and training for trainees. While trainees, although not fully trained, could offer a degree of service and attract funding, their training needs had a negative effect on the performance of their supervisors. Performance was measured through governmental targets such as

guaranteeing patients no more than a four-week wait for a consultation, and the need to avoid breaches of these deadlines in order to avoid fines or imposition of special measures (Coulson, 2009). These targets had led to management strategies like overbooking clinics, maximising scheduled operating lists and pulling trainees out of training to meet service needs. Participants were aware of examples of hospital managers adopting morally dubious strategies to manipulate results and avoid sanctions as illustrated in the quote below:

> '… the management will phone the patient saying, 'Mr X has got a really long waiting list, so do you want to have your operation done by Mr Y?' absolutely clear in the knowledge that Mr Y cannot do this operation, so the patient will say, 'Oh no, I'd rather stay with Mr X' so then the managers write, "Patient refused the date, take them off the waiting list" and the clock stops.' (P6)

Management strategies such as these were outside the control of consultant supervisors who nevertheless had to manage their practical effects. Participants spoke of the frustration of being unable to set up dedicated clinics focused on individual surgical conditions for the benefit of the training e.g. a hernia repair clinic. Consequently, training activities were fragmented which also affected the continuity of training relationships.

PERFORMATIVITY INFLUENCE: This term was first applied to educational policy by Ball (2008) to describe the effects of the introduction of systems of regulation which could alter the working practices of those working within them. Participants in the study alluded to performativity that existed within their workplaces and were displayed by the organisation, the system, the trainee and the supervisor. The first two, at organisational level had a strong influence on the second two at an individual level.

The impact of past medical malpractice (Lunt, 2008) on health organisations had the effect of greater regulation in terms of training, with a particular focus on improving efficiency, performance and safety. It had led to tighter regulation around funding, the imposition of stringent targets and the demand for evidence, and through evidence, accountability. Those responsible for delivering training at a national level responded by introducing electronic training management systems (e.g. to deliver the curriculum). These served as a vehicle for the introduction of key criteria aimed at promoting accountability and transparency and furnishing evidence for public reassurance. In their turn, these management systems employed assessment checklists and documented meetings which drove training. These developments tended to impose an additional administrative burden on supervisors, requiring them to complete online forms, meet trainees at set times and formally document all feedback. The participants complied with the demands of the system to varying degrees according to their ability to learn and manage what they felt was a complex administrative system.

> ‘the paperwork attached to it is not always the enjoyable part - digital, electronic paperwork, making sure you've recorded everything that you’re meant to record.’ (P2)

> ‘Making time for doing what we’ve got to do, so all the forms that need to be completed [are completed].’ (P3)

The key concern for trainees was thought to be the target number of assessments they were required to complete in order to be allowed to proceed to the next training level. Surgical trainees were encouraged to undertake as much operative work as possible, sometimes at the expense of attending clinics which were the main vehicle for discovering the needs of the patients on which they were operating.

WORKPLACE-BASED ASSESSMENT INFLUENCE: An aim of the curriculum developers was that WBA would enhance the direct observation of trainees – and thereby also provide a primary source of evidence on which to judge trainee suitability for greater entrustment. WBA was a series of forms with competencies matched to different training situations, as outlined in appendix 9. These were completed by trainees and supervisors to provide a judgement aid, prompt formative feedback and support the trainee-supervisor relationship. A crucial part of WBA was the verbal discussion that followed the observation, a stage in the process illustrated by the red dotted line in figure 6 below:

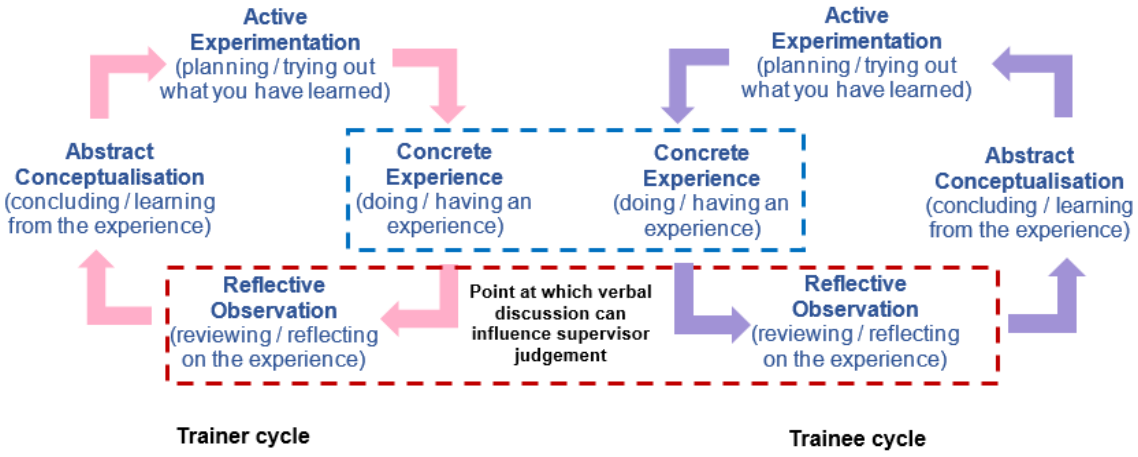

Figure 6: Trainee / supervisor model of learning (adapted from Kolb's cycle of experiential learning (1984)

The verbal discussion was intended to follow the common concrete experience, illustrated by the blue dotted line. Through discussion, the independent reflections for abstract conceptualisation of both trainee and supervisor should surface and be shared in order that an understanding and consensus might be reached. However, a key finding emerging from this study was that WBA was not used for formative discussion nor considered when making judgements. Participants tended to see WBA as a mandatory element to be tolerated rather than a formative experience where trainees were helped to learn. Having to

undertake WBA helped to provide opportunities for supervisors to observe trainee qualities but these were often not followed through to verbal feedback.

> "[In] clinics, very difficult to give feedback because my clinics tend to be slightly overbooked." (P1)
>
> [Do you give your trainees verbal feedback?] "To be honest, no, not really! I don't. Maybe I should do more, consultants know where you are in terms of whether you're progressing appropriately but no it's not something I do very much." (P3)

The implications of the lack of discussion were that supervisors risked losing trainees' views of what occurred and the consequent opportunity to place trainee actions and performance in context, possibly laying supervisors open to forming biased views of trainee performance. By missing the opportunity illustrated by the red dotted line in figure 6, understandings between supervisors and trainees risked divergence. Instead, supervisor tendencies appeared to reflect Shanteau's definition of the assessment expert with diagnostic expertise (1988). Rather than making decisions based on a large amount of assessment information, supervisors diagnosed the areas in which trainees needed development through the experience of having worked with them. The amount of time that trainees and supervisors spent together had, however, been constricted by curriculum requirements for formal documentation as well as the pressures of schedules on daily activity within hospitals. Consequently, supervisors could be left with quite a narrow picture of the dispositions of trainees.

TIME INFLUENCE: The third, and the most referenced, barrier was the time constriction placed upon training activity by the service and system changes outlined above. Participants felt that their overriding responsibility in guiding trainee development was their need to gauge trainees' levels of practical

capability e.g. how they constructed a treatment plan. To do this they needed to establish a training relationship which would allow them to better judge a trainee's personal level of development. Participants felt it was important that trainees were allowed the time necessary to learn skills thoroughly before they built up speed, they believed that this was the key to building not only skill but knowledge and confidence.

> "… you don't want people to go fast and feel that they are not in control and become slapdash because they're quicker but it's a combination I think of confidence and knowledge, but if you've got the knowledge you should develop more confidence. So, I think if people are operating slowly it could be that they have a gap in their knowledge of what they might encounter. This is a tricky one because we do all operate at different speeds … Some of it is practise and experience and you make sure you address any knowledge gaps." (P5)

While skills tended to develop through practice, trainees also needed to experience for themselves the breadth of variation through exposure to anomalies and rarer conditions. Supervisors' long-term aspirations for their trainees' development contrasted sharply with the short-term goals imposed by targets with which hospital managers were preoccupied. Consultants were pressured to complete the day's quota of operations, either sacrificing some training opportunities, intervening to save time, or devising creative ways of allowing trainees to become involved.

> "If you take time to teach a medical student how to stitch up a wound for the first time, you could spend 20 minutes rather than 10 minutes on it so what you might do is let them do a few stitches and then take over from them." (P5)

While hospitals provided a rich training ground within which trainees could learn and which corresponded closely with the curriculum theory of experiential

learning (Kolb, 1934; Dewey, 1916; Lewin, 1946; Piaget, 1964), trainees' and supervisors' ability to jointly exploit this landscape for training was constrained and shaped by the needs of the system and the targets imposed by management.

Level 2: Identification influences

In order to obtain a rounded picture of trainees and set their judgements in context, participants tended to draw on a number of other sources of information. The primary source other than direct observation was the views of other supervisors.

VIEWS OF OTHER SUPERVISORS: The participants tended to be involved in supervisor groups (as faculty communities) that collectively discussed trainee development and which also provided advanced information to the group about experiences with trainees. They would, for example, exchange emails to make other supervisors aware of important aspects of trainee behaviour or performance e.g. if a trainee had very poor technical skills, warning other supervisors to check their work very carefully.

> "When a new trainee joins, after a month, most of us will have an email from each other saying he's all right he can do this, this, this he cannot do this, this, this." (P1)

> "… if somebody says, "I'm just a bit worried about them," you're going to be less enthusiastic about letting them get on and do their own thing, you're going to spend a bit more time observing and making sure they are doing the right thing." (P3)

This tendency by the participants corresponded with expert behaviour according to Dreyfus and Dreyfus (2005). Experts tended to protect against a narrow outlook that might disregard possible alternatives by consulting others. In

Shanteau's view too, experts sought group feedback and were willing to make adjustments to their views (1988). According to Yanow and Tsoukas (2009), Schön emphasised the importance of a community of practitioners for providing its members with a common body of relevant knowledge. Faculty groups had the potential to provide collective understandings that situated supervisors relative to particular standards of practice. However, two of the participants mentioned that these resources were diminishing under time pressures.

Level 3: Internalisation influences

While the primary internalisation influence was direct observation of trainees, assisted by other supervisors, participants also tended to draw on grades, trainee views, junior doctor actions, nurses' views, patients and supervision levels.

DIRECT OBSERVATION: The direct observation of trainees as they carried out workplace tasks formed the primary source of evidence about trainee performance and progression and closely resembled the 'grounded' type of trust theorised by Ten Cate (2006).

> "Working regularly with someone, most of the time that's how it works, so that the same registrar does clinics with you, does the operating list with you, does the ward work with you allows you to have a much more natural progression …" (P2)

It tended to follow a pattern explained by Kolb's experiential learning cycle (1984), shown in figure 6. Working together, supervisor and trainee actively experienced a concrete training situation such as an operative procedure. The supervisor reflected on the experience as it happened, independently of the trainee's own reflection, leading to a personal abstract conceptualisation about the trainee's

ability. During the observation (and following it), the supervisor adjusted training according to the judgement made.

> "You very quickly get a sense of what you think this person can do and how sensible they are, and what they are confident doing, just from the way they position themselves, the way they assist you, the way they initially do simple things like close the skin …and then over time you let them do operations and hopefully you see them progressing." (P5)

Participants mentioned being able to trust quickly in relation to the more simple, operative tasks e.g. when they saw trainees positioning patients, in how they gave assistance and in making an incision. However, participants also felt they knew trainees well enough to extrapolate their knowledge of trainee abilities to areas which they had not seen first-hand as also encountered by Hauer *et al*. (2015).

> "I rely on them to keep the clinics running and see patients, and they come and see me if they don't understand something and so I'm accessible from that point of view but … they're seeing the simpler [cases]." (P5)

GRADES: Two participants (P1 and 5) thought that examination grades or previous training location were useful starting points for their judgements at the beginning of a placement but felt that it was important to keep an open mind. This view was consistent with Ten Cate's 'presumptive' type of trust (based solely on credentials, without prior interaction) in section 4.6.

TRAINEE VIEW: Some of the participants began by drawing on trainee views of their own ability and experience, a position more consistent with Ten Cate's 'initial' type of trust (based on first impressions).

> “Just tell me what you can do, what can you do?’ And they’ll say, ‘I can do a circumcision, I can do a hernia, but I need a hand with it, I’ve helped with appendices, but I’ve never done an appendix”. (P2)

> “The first time you meet someone and you ask them what they’ve done, you get a feel for what level they’re at … where he’d worked, who he’d worked for, that he’d done breast clinics and so I knew he’d be able to work alongside me in the breast clinic from when he arrived.” (P5)

However, as training progressed, they tended to assess trainees through their narratives about the work they had undertaken e.g. when they recounted to supervisors their conversations with patients, their planning in clinical consultations and operative situations and reasons for actions and decisions taken. Trainees’ descriptions might then have been compared with the consequent results e.g. operative outcome, choice and results of investigations and clear and informative clinic letters to GPs.

> “I see their letters, in the multi-disciplinary meetings I look at their note-keeping, so I get some feedback there about whether they’ve done it right.” (P5)

> “… he would come up and say, ‘We’ve got this patient, he’s one of yours, he’s had this problem, I’ve done this, is that ok?’ So, you just got to know what he was like, … and very quickly I feel rightly or wrongly an idea of what they’re like.” (P6)

All the participants mentioned that they used informal mechanisms such as discussions in the coffee room between cases as a means of gauging trainee values, personality, communication skills and professionalism as well as clinical ability rather than through formal assessments. They felt that talking was a powerful means of assessing trainees in a rounded and natural manner, so that trainees were unaware of being assessed.

JUNIOR DOCTOR ACTIONS: It was common practice in the health service for senior trainees to teach junior colleagues the straightforward aspects of the syllabus e.g. basic suturing skills. From the reactions of juniors, supervisors tended to gauge not only trainee teaching skills but also their level of competence and leadership Participants considered that it portended well when juniors who were struggling called upon more senior trainees rather than appealing to consultants.

> "the good ones, they usually get along with the juniors, so the juniors usually ring them first and if they're not that good or not committed they usually bypass them, then I'll know … if you suddenly realise that there is more of a direct communication without the registrar [senior trainee] then you know that something's missing there." (P1)

NURSES' VIEWS: Nurses and other healthcare professionals were also an important source of intelligence. They supported trainees in different environments and had a valuable perspective on how they approached day-to-day tasks when the consultant was absent. They tended to report to the consultant in charge when trainee performance was above or below the norm. Some of the participants expressly invited nursing staff to provide regular reports. Participants felt that a key problem area for trainees in difficulty was the relationship with support staff. Through nurses, supervisors could identify trainee technical skill level e.g. taking bloods; clinical decision-making e.g. about urgency of cases; and professional behaviours e.g. respect for team members.

> "… the theatre nurses have been told if the patient is taking longer than expected or if [the] senior [trainee is] in trouble or if the registrar [senior trainee] is reluctant to call us but is struggling we would like to know." (P1)

> "I found out via the nurses and the clinic letters and patients that actually his knowledge was poor …" (P5)

The importance of social relationships was a feature recognised in curriculum assessment, requiring trainees to demonstrate that they could form good working relationships with the multi-professional team (MPT) in order to deliver clinical care effectively. The MPT had an assessment role, most importantly through the MSF assessment (in which trainees' self-assessments about their own professional behaviour was compared with the views of their team members, see appendix 9). While participants highly valued MPT perspectives about trainees, they were aware of the existence of potential biases caused by cultural differences. For example, when trainees from different ethnic groups began a new placement, there was a high possibility that the differences in the way people communicated could give rise to misunderstandings as in the example below.

> "… this same nurse who was a bit awkward complained to me about this same doctor and said he was rude to her and I'd happened to be sitting next to him when he took the call …They [the nurses] wanted a meal break and he was saying this patient needs to be done and you can have a quick cup of tea if it takes five minutes but we can't afford to leave this patient half an hour or an hour, and I heard him and he just said that pleasantly but firmly, but he wasn't saying what they [the nurses] wanted him to say and that was being counted as rudeness. He was also a doctor from overseas and I think, sadly, they do sometimes get a harder time." (P2)

Surgical training required trainees to rotate through relatively short placements in different units. Consequently, supporting staff had a short time in which to get to know individual trainees. Nurses were also likely to be subject to the same tensions between service to the patient and helping to support training with a subsequent impact on working relationships.

PATIENTS: Participants also took account of patient views and outcomes. As well as patient complaints and compliments, supervisors were aware if trainees asked

patients to return to clinics unnecessarily and whether patients were left uncertain or dissatisfied after their consultations with trainees.

> "[patients] would then phone … and say that what he said didn't make sense or, 'This guy was rude to me' … they are quite used to seeing somebody different, but they know what to expect." (P5)

SUPERVISION AS A SCAFFOLD: The manner in which participants supervised trainee learning was consistent with the supervision framework governed by the curriculum as an educational 'scaffolding' teaching strategy (Van Der Stuyf, 2002). It corresponded with the principle that the overall responsibility for managing patient care rested with consultant supervisors. The scaffolding of learning began with trainees observing supervisors, progressing to assisting supervisors and finally, supervisors assisting trainees. Trainees advanced from managing straightforward cases independently to more complex cases with access to help from supervisors as needed. The level of supervision decreased, and the level of case complexity increased as trainees became proficient in technical and non-technical skills and professional judgement. The scaffolding nature of supervision corresponded with the proposed supervision levels in table 2 (page 53) and meant that supervisors could allow trainees to do parts of activities with help and support, removing the supports or pieces of the scaffold as trainees were able to do more on their own until they could perform without help. Participants also described the work as modular, conducive to being broken down into self-contained units which could be taught and assessed, learned and practised in a deliberative manner.

> "… let them do bits of operations to start with and … also because there are some parts that are more straightforward than others and then they move up to doing the whole procedure so the most juniors practise

> their suturing, and practise closing the wound … and then move on to more complex things. When you remove part of a bowel for example you have to mobilise it, remove it from the things that are holding it in place and divide the blood vessels that are supplying it and then connect two ends together, they are all separate bits that people might do as they get experienced." (P5)

Participants encouraged trainees to perform to a level just beyond their competence when in the presence of the supervisor because the supervisor could take over at any point if the trainee was struggling. For example, participants described carrying out clinics with trainees independently but in adjacent rooms. Before allowing trainees to perform, supervisors were likely to discuss the steps with trainees and probe them. Supervisors had the option of selecting appropriate cases, setting out the method and parameters of each task. These included the safest method, materials and specific technique consistent with both the trainee's and the supervisor's expertise.

> "… you have to get them to talk through what they are doing before they do it or replay what they see you doing before they start it so that you know they've got the right steps in mind you might let them draw where the incision is going to be and that will illustrate if they've thought about what their approach is going to be." (P5)

The scaffolding framework, originating from Vygotsky's sociocultural theory included the concept of the *zone of proximal* development (ZPD), (Vygotsky, 1987; Bruner, 1984; Wertsch, 1984). The ZPD was the distance between what a learner could do by themselves and the next learning that they could be helped to achieve with competent assistance. Bruner (op. cit.) referred to it as 'mental sharing' where those who 'know' more share it with those who know less (p. 94). As a supervision strategy, scaffolding provided individualised support based on

the learner's ZPD and suggested that supervisors made judgements as they reduced the scaffolding.

> "… if you're going to do a long and difficult operation, then for a trainee who is on the limit of their ability on that operation, rather than trying to get them to do it from the beginning … say, 'I'll get you to that bit and then you do that bit and then next time you'll do that bit and something else' …" (P2)

Therefore, supervisors had a clear framework for applying levels of supervision from merely observing to performing complex cases independently. The nature of modular learning assisted this process. However, rather than using a curriculum process for judging trainee competence, participants had created their own tacit and intuitive approaches to judging trainee competence.

7.2 Decision-making

This section describes the nature and quality of surgical decision-making and the role of trust as the critical factor in deciding whether trainees should be allowed to perform autonomously. It goes on to depict the determinants of supervisor trust in trainees, how trust tended to be granted and what made trusting more likely. It also sets out the strategies that enabled trainees to perform alone for the first time as part of the incremental process leading to independent practice.

Decision-making was referred to by all participants as the single most important element of their identity as surgeons, it was the most critical element of trainee performance judged by supervisors. Decision-making was closely associated with clinical judgement which transcended practical skills.

> “If he can make the decisions the rest of it will fall into place. Because you can have the best hands it the world, if you make the wrong decisions, you’re not going to make it.” (P6)

The quality of clinical decisions

Participants referred to the characteristics of good quality decision-making in surgery as the ability to make justifiable and confident decisions quickly in uncertain situations. Important to this was the ability to synthesise and prioritise information which may be limited, to have situational awareness of the likely impact of the decision on others such as patients and colleagues but also on resources and the service. The quality of trainee decisions could be ascertained through the appropriateness of patient management plans, in how and when they called upon the consultant for help and how they summarised cases (patient status, judgements made, and actions taken) such as during handover. Trainees were expected to use professional judgement through applied knowledge to make decisions appropriate to their stage of training. Essential also was the capacity for insight into their limitations and understanding about when to seek advice. At the start of training, trainees were expected to be able to make straightforward treatment decisions. As training progressed and their knowledge and experience increased, they were required to make more complex decisions in uncertain situations.

> “… the case with technical things, that’s relatively easy, cos you can say someone at this stage would usually be able to do this operation on their own and this person has to do with help. If it's to do with decision-making that’s much harder … he just didn’t make good plans, and it wasn't because he was lazy, he didn’t need to try harder, he just didn't seem to see it”. (P2)

Surgical choices tended not to be clear cut and decision-making needed to balance the risks of surgery against the risks of doing nothing. Participant 6 for example, referred to the difficulty of making decisions where refusing operative care meant ending someone's life.

> "So whether they [the trainee] just didn't get it, it didn't make sense to them, or whether they didn't want to make the decision where the answer was no, because if a neurosurgeon says no, most referring departments will think, 'the patients going to die, so we'll do nothing,' which is not what we said, what we said was that there was no neurosurgical intervention, that often gets interpreted as let's put him in a side room and let him die …" (P6)

The approach by Coles and Fish (1998) was allied with the participants' view that even at the early stages of training, trainees were expected to be more than the 'novice' suggested by Dreyfus and Dreyfus (2005). Trainees were expected to apply their level of knowledge to the real world without following context-free rules or a rigid approach. At each stage of training, trainees were expected to have a personal investment (e.g. as characterised by 'competent performer' which was two stages higher than the novice in the Dreyfus model). While the Dreyfus model showed staged progression, the Coles and Fish model corresponded better with a continuum and was situationally dependent. As complexity and uncertainty increased, the clinical decisions called upon demanded practical wisdom which corresponded with true clinical judgement.

> "… from a ward round point of view I would actually say that [the junior trainee] would rank higher than [the senior trainee] in the overall management of a ward round, but from a knowledge base of individual patients [the senior trainee] would be better … it's based on knowledge and things, the more senior a trainee becomes the less they know about each patient on the ward, but the more they'd know about the condition and, therefore, how you would manage each of the patients". (P6)

The speed of decisions

Participants expected trainees to make quick, justifiable and confident decisions which they felt distinguished surgery from other medical disciplines.

> "… we get called at night, five, ten, twenty times a night on just really sparse information and you have to live with it, you have to live with the decision you are making … physicians also make decisions, but it's a very different way, it's usually made on a whole bunch of tests and then we'll decide, whereas we're much more likely to make a snap decision and that's it." (P6)

Participants felt that confidence was connected to knowledge and the ability to make quick and justifiable decisions, which agreed with Croskerry, Petrie, Reilly and Tait's (2014) view that the clinicians who were more knowledgeable were likely to be confident and make decisions quickly. In contrast, they observed that those who were not confident were more likely to take longer to make decisions.

> "… it's a combination I think of confidence and knowledge, but if you've got the knowledge you should develop more confidence. I think if people are operating slowly it could be that they have a gap in their knowledge of what they might encounter." (P5)

> "… you can only make decisions based on knowledge so, the senior trainee has a lot more knowledge and experience than the junior trainee, so is happy to make decisions." (P6)

The nature of trust in decision-making

As explained in detail in sections 4.3 to 4.6, trust and judgement were linked because the notion of trust was an influencing factor in supervisor decisions about the level of supervision trainees required. The notion of trust was central to the concept of supervision because without trust, trainees could not progress. Trust

was, therefore, considered to offer a more meaningful and intuitive taxonomy for practice than a list of competencies.

> "So, the big thing is trust, if they lie to you they're buggered ok? … we all make mistakes, but don't ever lie to me about it … because if I stop trusting you, to a large extent, it's over, I don't trust you again and then your training's buggered because every time I have a list with a really good case, you're not the one I'm calling … I'm going to be coming in to see what you're doing, I'm going to be overseeing your operations, you're not going to get the freedom to fly." (P6)

As suggested by Govaerts (2017), entrustment (allied to trust and defined in section 2.1 as *the granting of permission to perform a function associated with patient safety*) was likely to be temporal, person, task, and context dependent, influenced by factors related to the trustor, their experiences and working relationships as well as influenced by political, financial and cultural factors. Grounded trust (Ten Cate *et al*., 2016), which was based on intensive contact with the trainee, was customary through the vehicle of the training placement and in the clinical context there was a connection between trust and the amount of time it took to develop. Trusting was linked with the quality of the supervisor-trainee relationship. Trust developed more quickly when participants knew trainees well, liked and understood them.

> "I like to know about people that I'm working with … you ask them what they've done before, what type of surgeon they want to be, where they're up to with their exams and I think I've always asked those things fairly early on with people cos if I'm working with someone I want to know immediately what to expect of them." (P5)

Some felt that training relationships took longer to establish because of the nature of new shift patterns and consequently it took longer to trust trainees to carry out larger, more holistic responsibilities. Supervisors were more likely to trust trainees

when they felt confident that they had sufficient knowledge for the task and could apply it appropriately. Their trust would extend for example from simple mastectomy to breast reconstruction.

> " … a simple mastectomy means you're removing a lot of the breast and a lot of the skin just to leave the chest wall flat, but if you're going to do a reconstruction you have to leave a lot of that skin and it's important that the skin you do leave has the right amount of fat under it, there's less room for getting into the wrong place and damaging the skin …, 'cos it's important that the skin is saved and is healthy. So, you have to be able to get the right plane … getting the right layer between tissues, use the right instruments, use your assistant [senior trainee assisted by junior trainee], so you know when you haven't got a good view of what you're doing, know whether it's because you're standing in the wrong place or your assistant's not pulling in the right direction, or your light isn't right, and those sort of things are more complex." (P5)

Trust was demonstrated through the granting of degrees of freedom to perform tasks (like scaffolding mentioned in chapter 7, page 135) such as part or whole operative procedures or management of a patient case with the supervisor directing at one extreme or delegating at the other. Participants were asked how they decided that a trainee could be trusted with an unsupervised task for the first time. They said that they were more likely to be able to trust if trainees had been immersed in the patient case from the beginning. They felt this indicated a degree of commitment because there was a lot of pressure on trainees from the service to solely perform operations. It also indicated continuity with the patient; understanding of patient history, preferences, risks and objectives. Similarly, it gave trainee and supervisor a trajectory of development upon which to build grounded trust (op. cit.) and ipsative assessment (Hughes, 2014). Another significant element of the ability to trust was linked with trainees' insight into their own limitations and their ability to seek help, which was comparable to judicious

decision-making. Trainees were more likely to be trusted if they demonstrated a tendency to call for advice and help when they encountered a situation outside their sphere of competence. Trainees were expected to be honest about what they did and did not know and what they had done wrongly. Supervisors felt that if trainees demonstrated honesty, they could be trusted first with simple responsibilities and then more complex tasks. The participants, as consultants, took ownership of the care of patients listed under their name. As such, they were prepared to be woken in the middle of the night when 'on call'.

> "If you think that you might need to call me, then you have to call me … And if you're not sure and you want to run something by me, then you call me, the patients are mine, they're not yours … and it's my name at the end of the day ...". (P6)

> "[He] knows what he needs to let you know about … so if he calls me at night he'll tell me what he wants, which I find very helpful particularly if you're woken and you need to know ..." (P2)

Naturally, they expected trainees to make sensible decisions and call them only when it was necessary and when they were clear about what advice they needed.

> "… we did have a trainee like that, really good at ST1 and 2, came to registrar level at ST3, couldn't make a decision to save his life, it was awful being on-call with him, it was just a disaster, 'Oh shit, I'm going to get no sleep tonight, I'm going to get lots of phone calls, no decisions' …" (P6)

Trusting for the first time

Trainees who were deemed high-fliers would be trusted with more complex procedures more quickly. However, participants also trusted for the first time in order to motivate trainees, to stimulate their interest and encourage them to do better.

> "… even though it might be very complicated, try and get them to do something … part of that is because I don't want them to get bored, because if you get bored you don't concentrate on what we're doing." (P6)

The participants felt themselves to be under continuous intense pressure from service needs, consequently this could also be a deciding factor in whether to let trainees perform alone. This, however, did not mean that they were totally unsupervised because other team members (e.g. senior theatre nurses) were always likely to be involved and able to offer secondary oversight. Alternatively, the supervisor could be within calling distance in a room nearby. To help them trust for the first time, participants had at their disposal a number of management techniques and controls. Chief amongst these was the ability to negotiate and plan the task with trainees in advance. They might first talk through how trainees would undertake the task, questioning the trainee's knowledge, method and approach. They might talk through risks such as uncontrollable bleeding and how to mitigate these and what to do if unforeseen circumstances arose (such as complications or a critical incident). The aim was that trusting for the first time was a result of a gradual and incremental process in which trainees had proven their knowledge, skills and judgement so that removing the final level of support was a natural extension.

In summary, while the environment in which trainees were situated had a significant impact on both how they were judged and on what they were judged, supervisors had some freedom to apply resources creatively. They drew on frameworks for supervision and a number of other sources of evidence on which to base their judgements. These judgements were linked to supervisor confidence in trainees, leading to decisions to trust demonstrated through

degrees of independence. The participants regarded the ability to make decisions quickly as crucial to the character of surgery, it was also the most vital component of supervisor judgement and closely associated with clinical judgement.

The supervisor judgement process tended to be unconscious and tacit, spontaneous and intuitive, linked to the actions of everyday life. However, a more explicit pattern of judgement theorised as being exercised by supervisors is described in the next chapter.

Chapter 8: Judgement in Action

> "We can recognize and describe deviations from a norm very much more clearly than we can describe the norm itself"
>
> Vickers, 1978 (cited in Schön, p. 53)

Participants in the study were able to recognise when trainees were or were not performing or progressing satisfactorily but were unaware of their own intuitive frame or cognitive process of judgement. A new perspective, emerging from this study, provides and makes explicit a theory about the intuitive processes and pattern of judgement involved and suggests how it could be practically applied.

The theory of Judgement in Action, which emerged through this study as an adaptation of Schön's theory of reflection in action, lends a new and incisive frame of reference for how surgical supervisors think when making judgement decisions.

Judgement in Action involved a simultaneous interaction between clinical judgement and trainee assessment. In other words, supervisors used *their* clinical judgement, while working with trainees, to judge the quality of *trainee* clinical judgements about patient care. Judgement and decision-making were, therefore, key components of successful surgical teaching and practice.

Judgement in Action

Participants revealed that when they judged performance, they drew heavily on the qualities they valued in trainees. Their approach was based on four types of

capability sub-theme: Competence, Decision-making Administration and Relationships. These served as sub-themes and within each were qualities that trainees were expected to exhibit in order to be considered as progressing satisfactorily. Examples of these qualities under each sub-theme were as follows:

COMPETENCE SUB-THEME

The acquisition of knowledge and skill:

- Knowledge
- Skill and technical expertise to perform tasks independently
- Timeliness
- Calmness / confidence
- Professionalism / probity

Participants described knowledge, skills and attitudes as underpinning surgical ability. They felt that these areas were generally easy to learn through practise and to demonstrate through individual tasks which were episodic in nature. Competence was immediately obvious and generally easy to assess in an objective way e.g. dexterous hand movements, correct use of surgical instruments, positioning of the patient and ability to assist. However, when viewed in isolation it was not considered sufficient for high-level practice.

DECISION-MAKING SUB-THEME

The capacity to make a logical and justifiable choice from available options quickly:

- Ability to make decisions appropriate for the training stage
- Ability to make justifiable and confident decisions quickly

- Ability to synthesise and prioritise
- Ability to deal with complexity and insufficient information
- Ability to have an awareness of the impact of decisions
- Ability to have insight into one's own ability and know when to call for help

The participants perceived the single most important attribute to successful surgery as the ability to make justifiable and rapid clinical decisions based on knowledge and experience. Clinical decisions dictated the treatment path and all actions that followed and were closely associated with the surgeon as leader of the clinical team. Because it was referred to by all the participants on numerous occasions, it emerged as a pre-eminent sub-theme. The decision-making sub-theme was closely related to Coles' assertion that caring professionals required phronesis (practical wisdom) as outlined in chapter 4. They were not only technically accountable for their actions but also morally answerable (2002). As argued by Croskerry, Petrie and Tait (2014), this area was likely to involve both *intuitive* decisions (which were fast with minimal cognitive resource) and *analytical* decisions (slow, deliberate and conscious). While both intuitive and analytical decision-making were important, surgical decisions tended to need to be made quickly. However, practitioner employment of both had to be apposite.

ADMINISTRATION SUB-THEME

Methodical and organisational capacity in the execution of a task:

- Logistical
- Proactive
- Responsible
- Committed

This area related to the trainee's ability to organise complex activities such as out-patient clinics and operating lists. It involved the trainee working on their own initiative, planning ahead, taking responsibility for leading and managing people across different teams, managing time, resolving problems and keeping patients happy. These activities required holistic capabilities enacted over a protracted period of time rather than competence in individual tasks on single occasions. These skills were purported to be very highly valued by supervisors who saw this area of work as inconvenient and tiresome. It, therefore, lightened their workload and freed them up to focus on the areas where their expertise could be best manifested. Participants felt that trainee ability in this area made a major difference to their day, and their mood. Through these activities, they came to believe in whether trainees were performing well, using words such as 'reliable', 'committed' and 'responsible'. Good administrative ability helped to improve the relationship between trainees and supervisors and was likely to lead to better training opportunities for trainees. Administrative ability was perceived as bureaucratic but nevertheless important because of the potential it had for alleviating the burden on the supervisor.

> "I've got a problem, I need this sorted, ok" and it's done, I don't need to ask him again. If for some reason he can't do it he will come and find me and say, "I tried to speak to the patient, tried the phone number, tried the GP, can't get hold of them, don't know where they are, so I've done a letter".' (P6)

RELATIONSHIPS SUB-THEME

This area related to the notion of emotional intelligence, defined as 'the effective regulation of emotion in self and others, and the use of feelings to motivate, plan, and achieve in one's life' (Salovey & Mayer 1990, p. 185).

- With supervisors
- With juniors
- With patients
- With nurses
- With the surgical team
- Social skills (including appropriate competitiveness and cultural fit)

Because the surgical treatment of patients involved multi-disciplinary teams, surgeons were required to work closely with allied healthcare professionals such as theatre nurses and other medical doctors such as anaesthetists. In daily practice, different groups had different underlying intentions and motives and the potential for conflict leading to friction and disharmony was high. Colleagues were more inclined to facilitate a surgeon's needs if s/he was perceived to be respectful (e.g. attentive to the role of the other person), empathetic (about their needs) and facilitative (helping in some concrete way). Persaud (2011) suggested that social rewards were mediated through actions such as these and, for trainees, this could lead to benefits such as favourable assessments, access to resources and inclusivity. The social skills that trainees needed in order to enhance relationships and team-working could, therefore, significantly influence the environment in which they worked, helping to overcome the perceived barriers to surgical practice and helping to convince supervisors of their suitability for greater entrustment. However, at the same time, they had to be appropriately competitive in order to obtain the opportunities they needed for their own training. The ability to manage relationships was at the heart of professionalism.

These sub-themes corresponded with qualities depicted by different writers in chapter 4. Judgement based on trainee qualities was one of four factors emphasised in judgement by Sterkenburg et al. (2010). These included notions of benevolence, integrity, recognition of one's own limitations, self-efficacy, conscientiousness and honesty. The competence and decision-making sub-themes corresponded closely with Shanteau's (1988) work that highlighted the requirement for experts to have *competence*, *mental processes* and *cognition*. Relationships correspond with Shanteau's depiction of 'assessment experts' who had perceptual, diagnostic and action expertise as well as good social skills, while Hauer *et. al.* (2015) mentioned organisational skills, knowledge sharing and good reasoning.

These four sub-themes represented distinct types of capability and they could be differentiated according to two categories that set out whether they were:

1. ***Core*** (central) or ***Complementary*** to surgical practice and
2. ***Learning How*** (learned through instruction) or ***Learning Via*** (learned through experience and deliberative practice).

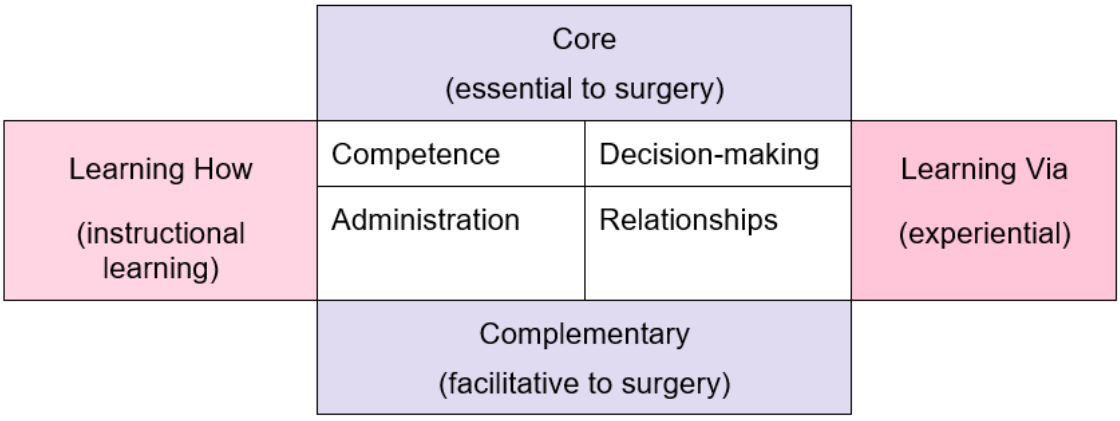

Figure 7: The four capabilities (sub-themes) of Judgement in Action

Core	Complementary
The Core category was at the heart of surgical practice. In isolation it lent itself to hierarchical settings in which all supporting elements were entirely subordinate. However, in modern surgical practice, there was a heavy reliance on contributory work carried out by auxiliary teams. A trainee surgeon who was technically competent and could make good decisions about patient care would meet the standards of this category but might lack the ability to engage other people e.g. because of rudeness to nursing staff and consequent inability to gain their commitment.	The Complementary category facilitated surgical activities by positively influencing elements of the environment, allowing surgical practice to take place. A trainee surgeon who had good organisational and social skills could successfully organise activities and gain the co-operation of colleagues. However, without core surgical skills they would be unable to perform as a surgeon.
Learning How	**Learning Via**
The Learning How category represented the basic elements of ability which could be learned through instruction and study; involving knowledge, understanding, practice and organisation. It provided the underlying structure on which higher functions sat. A trainee who had learned how to perform an operative procedure, to set up the theatre suite and organise the pre-operative team briefing would not necessarily be able to galvanise the team, lead and make appropriate treatment decisions.	The Learning Via category represented learning through experience and deliberate practice. It was a creative and transformational process in which existing competencies were adapted and tuned to new circumstances (adapted from Fraser & Greenhalgh, 2001). Learning Via also occurred on a more personal level akin to Judgement in Action where judgement occurred in the midst of an experience, allowing improvisation to benefit the situation.

Table 8: Categories of Core / Complementary and Learning How / Learning Via

Using these categorisations, the capability described as ‘competence’ was categorised as *Core-Learning How* because it was essential to surgical practice and tended to be learned through instruction. The capability described as ‘decision-making’ was categorised as *Core-Learning Via* because it was essential to surgical training but tended to be learned through experience. To be considered as performing satisfactorily, all four capabilities were important. However, the two categories of *Learning How* and *Learning Via* were not at an equal cognitive level. While both *Learning How* and *Learning Via* were essential to progression, *Learning Via* was of a higher order because it required a high degree of reasoning. As argued by Fraser and Greenhalgh (2001), this type of capability ‘could not be taught or passively assimilated as it was achieved through a transformation process in which existing competencies were adapted and tuned to new circumstances’ (p. 800). In this way, the cognitive processing involved could be described as creative behaviour. Learning Via also corresponded with the notion of deliberative judgement by Coles and Fish (1998) which had to involve considerable practise, reflection and analysis gained from learning from experience.

The capabilities and two categories are shown as quadrants in the graphical framework in figure 8 below. Trainees performing highly in all quadrants might be depicted as strong in all four areas. From this illustrative perspective, performing satisfactorily, could appear as an even red circle.

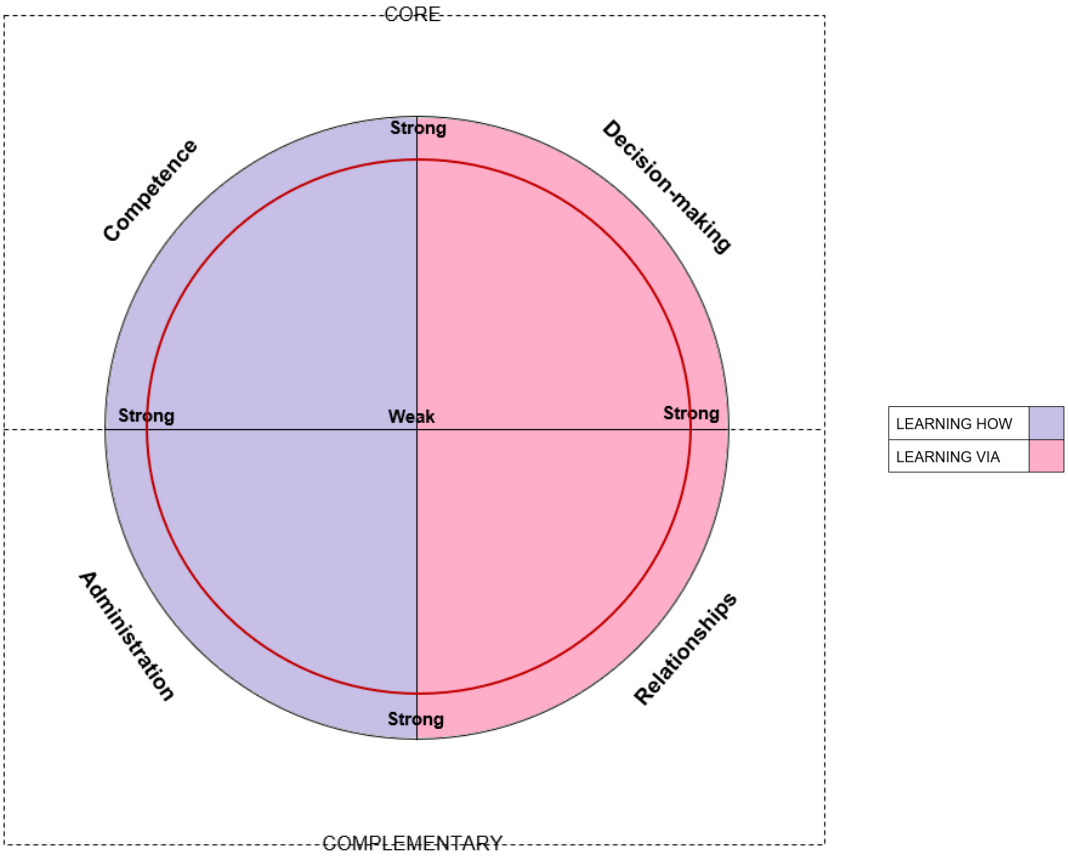

Figure 8: Supervisor judgement framework

Capability examples

A trainee who was competent and confident teaching juniors to clerk patients and take bloods but was reluctant to undertake operative work might be performing highly in administrative and relationship capabilities but underperforming for the stage of training in both competence and decision-making.

> " … she's very good at ward work almost still making that transition from FY2 or an ST1 to being an ST3, she will go off and very helpfully bleed the children, and make sure the paperwork's done, do some of the FY1s paperwork … but at the expense of operating time or ward rounds, so it's very helpful from the point of view of getting the work done, or team work, but actually she's missing out on some of the training actually doing some of the things that someone else is now supposed to do." (P4)

The judgement of this trainee's progress might be depicted as in figure 9:

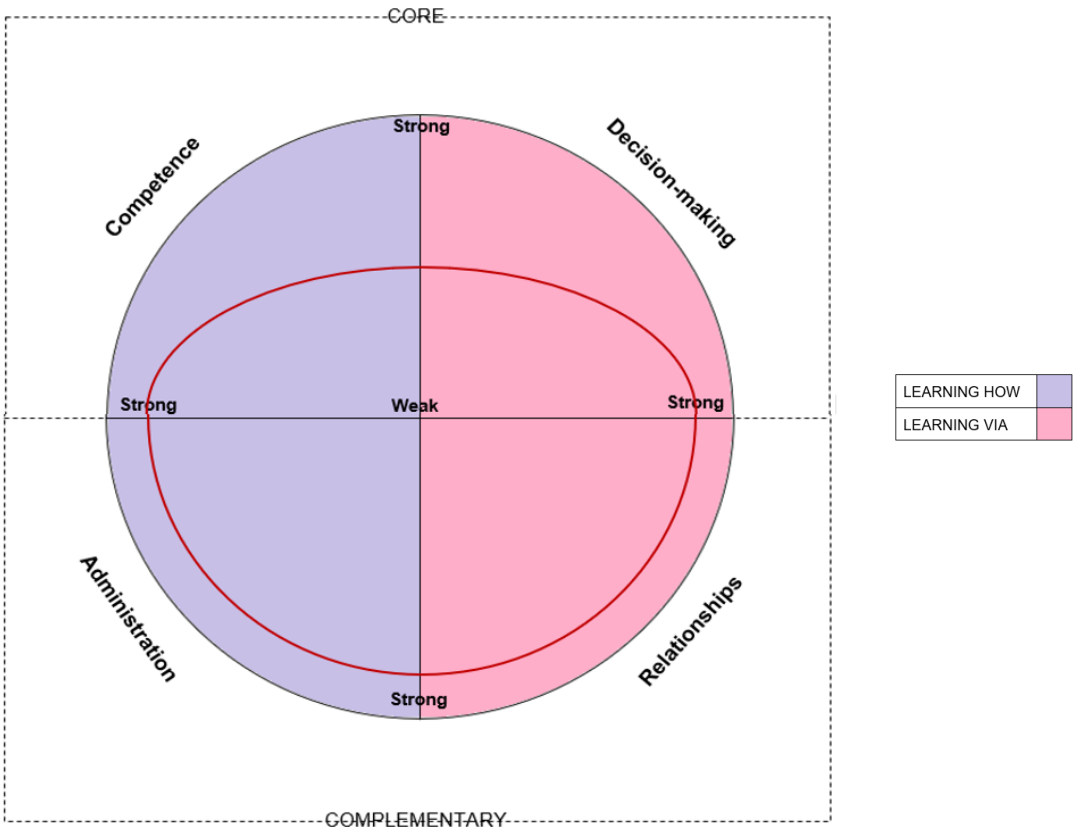

Figure 9: Example 1 judgement of trainee performance

Similarly, a trainee who had good operative and decision-making skills but did not communicate well with the theatre team would be performing highly in competence and decision-making but underperforming in relationships and likely to be also underperforming in administration because of the team-working dynamic.

> "I think he's very close as far as technical skills and decision-making … but he's perhaps not very good at dealing with people being awkward with him … I found when he was working with me that I was very happy with his work, but I think you can't let your work vary too much just because you do or don't get on with somebody." (P2)

The judgement of this trainee's progress might be depicted as in figure 10:

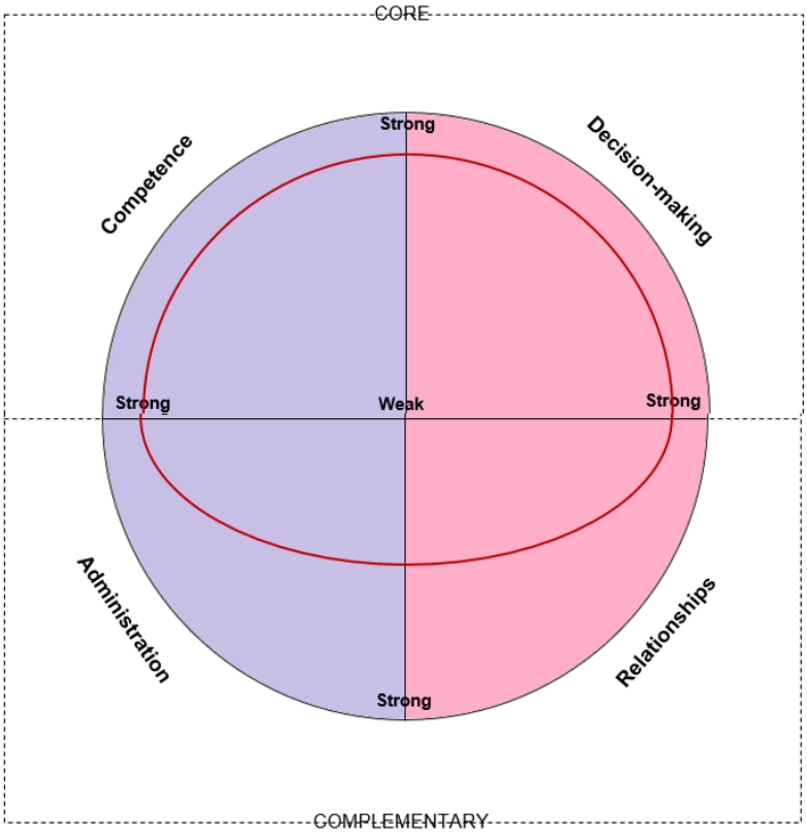

Figure 10: Example 2 judgement of trainee performance

Therefore, optimally successful trainees would be able to balance the four capabilities within the two categories, while trainees lacking in any area would be described as weak.

The study findings showed that supervisors tended to judge trainees more roundly than could be achieved through WBA. Their unconscious and spontaneous judgements were concurrent with the actions of everyday life – judging in action. The theory of Judgement in Action developed from evidence which suggested that supervisors tended to know more than they could express in words. The literature supported the link between the concepts of expertise, judgement and trust and their highly complex and domain-specific nature. Judgement in Action was a framework that allowed the adoption of these

concepts as a basis for formal assessment, designed around an understating of how supervisors tended to think. By adapting Schön's framework of reflection in action to the context of surgery, Judgement in Action, allowed a more contextualised description of supervisors' intuitive processes and helped to define the critical qualities of knowledge, experience and deliberate practice that would be requisite for anyone undertaking the role of supervisor.

The definition and process of Judgement in Action

Supervisors might be involved with two interconnected streams of reflection in action. As supervisors, their reflection in action related to judging trainees for trainee learning and assessment. As clinicians, their reflection in action related to judging the trainee's treatment of the patient for a healthy outcome. For supervisors in surgical training, both streams happened at the same time at the point of delivery of care. Because the main focus of supervisors was on making judgement decisions about the trainee's contribution while simultaneously making improvements to their own judgement practice, the term 'reflection' was subtly adapted to 'judgement' in action.

Unlike Schön's definition of reflection *on* action, which referred to looking back on actions after the event and considering what could be changed, Schön's definition of reflection *in* action was that it took place in the midst of action when learning and improvement occurred contemporaneously. In medicine (and especially in surgery), this also needed to unfold quickly as recognised by Alvesson, Hardy and Harley (2008) and Eraut (1994). Schön defined reflection in action as 'knowing in action' akin to knowing more than could be expressed in words, where 'innumerable judgements' were dependent on 'tacit recognitions, judgements and

skilful performances' (1990 p. 50). These kinds of judgements were only made known by a decision or action e.g. the supervisor's judgement about whether the size and position of an incision made by a trainee between two major organs meant that the trainee was ready to perform independently. Supervisors tended to gain an unconscious understanding of trainee ability through observing trainee fluencies. Such fluencies included trainee *technical abilities* (e.g. incising, positioning and suturing); *behavioural abilities* (e.g. listening, speaking, instructing); and *judgement* (e.g. improvising, prioritising and delegating). As a result, supervisors could be characterised as making judgements about trainee aptitude to solve patient problems corresponding with Shanteau's (1988) view of supervisors as 'assessment experts' managing uncertainty (page 65).

Schön's theory of reflection in action applied well to surgical practice and helped to take account of the swampy nature of the environment and to demonstrate how tacit judgement might work. In keeping with the metaphor of high ground and swamp, Schön defined reflection in action as taking place in the 'indeterminate zones' of the swampy lowland as practitioners went about 'the spontaneous, intuitive performance of the actions of everyday life' and showed themselves 'to be knowledgeable in a special way' (op. cit., p. 49). It also drew on concepts of praxis, or practical wisdom (Coles, 2002), as mentioned in section 4.4. According to Harris (1989), writing for the Association of Supervision and Curriculum Development, reflection in action was essential for dealing with problems that did not yield to technical instrumental solutions. Like Dewey (1916), Schön made a strong link between personal experience and learning as a simultaneous interaction. Dewey posited that all genuine education came about through experience (1986, p. 247).

Critiques of Schön's theory of reflection in action, suggested that he presented a dichotomy between technical rationality and personal experience, a dualistic rather than pluralistic approach (Harris, 1989). In contrast with this view, reflection in action as applied to surgery, could be seen as validating both the technical and experiential elements of judgement, a view supported by Kinsella (2007) and Eraut (1995). Judgement in Action, however, reflected the participant view that while technical competence and experiential learning were both important, the latter was primarily used in the judgement process and was characterised by uncertainty and value conflict.

Adapting Schön's framework of reflection in action to Judgement in Action involved four steps which corresponded well with the learning cycle developed by Kolb as shown in blue text in figure 11 on the next page. Judgement in Action characterised supervisors as observing trainee actions as part of their routine activity. Their judgement began when a routine training activity produced trainee performance that yielded unexpected results either positive or negative (which Schön expressed as a 'surprise', p. 68) and which might quicken the supervisor's interest. As an example, if a junior trainee in theatre, facing a previously unencountered complication, demonstrated s/he had recognised the implications and acted accordingly (by stopping, communicating with the team and making appropriate adjustments to instruments and pace), the supervisor, as a result of the positive surprise, might need to improvise or adapt on the spot in a unique manner (e.g. step back rather than intervene), leading to a revised opinion of the trainee's capability and alteration of the training plan, granting a higher level of trust in the future. According to Harris (1989), surprises interrupted the normal flow and signalled a lack of fit between existing knowing-in-action and the

situation at hand. In each instance, there were different implications for how practitioners might respond.

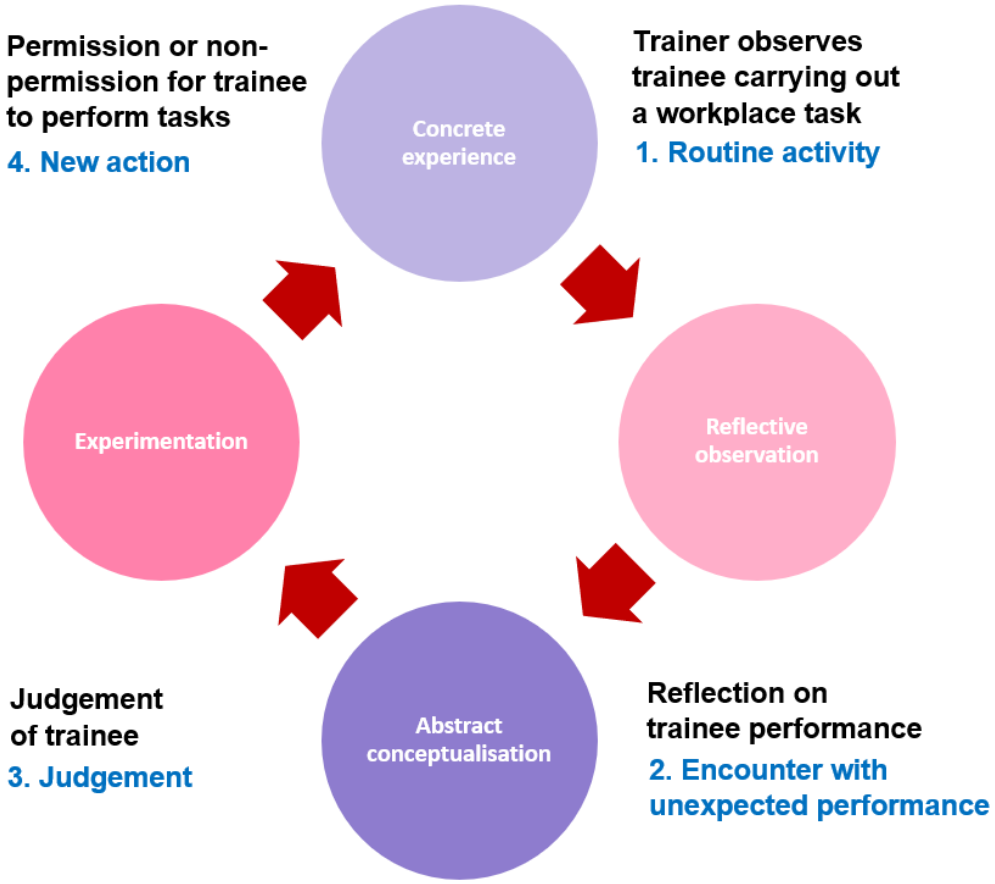

Figure11: Kolb cycle with reflection on action overlaid in blue

Another element of supervisor judgement, mentioned by participants, was their sense of responsibility and ownership of patient care. Treatment was administered in the supervisor's name and patient outcomes were subject to formal reporting. In this way and in confluence with Yanow and Tsoukas (2009), drawing on Dreyfus and Dreyfus (2005, pp. 784–5) and Schatzki (2005, pp. 471–2), surprises posed not only a mental challenge for supervisors but also an emotional one in 'situations that matter'. Harris suggested that improvisation was not without preparation as in music and theatre where performers used pre-established and rehearsed repertoires of situation-specific moves. He also

suggested that while some cases of improvisation might have appeared to be unique, they all necessarily drew on collective codes of practice and professional norms. The example quoted below might suggest that supervisors had a repertoire of judgements as they reduced their scaffolding.

> "… let her do it but with less supervision, put her slightly more under pressure see if she can make the right decisions and see whether actually the improvement is due to her ability or her personality, some people can do things when they are very protected and if you let them a little protection off, they may not be as good as that'." (P1)

According to Harris, those judging were thought to 'engage in a sort of dialogue' with the surprise and their repertoire of moves, choosing between possible answers which could offer solutions. These solutions were by their nature 'in the moment' and represented a non-cognitive orientation.

A limitation to Judgement in Action was the risk that supervisors could be drawn into patterns of error because of outdated judgements. To varying degrees, the participants felt that trainees quite often adopted a different approach to their training than had supervisors when they were trainees themselves. Trainees tended to want a more equal balance between their personal and work lives and consequently could tend to appear less vocational.

> "this guy was out of the theatre like a rabbit up a drainpipe basically, as soon as the last stitch was in, he never hung around to find out if anything else needed doing or would never even ask if it was alright to go, or if there was anything else you would want me to do, he just leaves". [Researcher: And where was he going?] "Home! I think, he was just getting out as soon as possible … he's in a slightly different situation, he sees it as a job that he comes and does" (P4)

Trainees tended to take breaks in their training pathways and seek experiences outside the training programme e.g. to work in low income countries. The ability of trainees might, therefore, not have been as accurately categorised by the older generation of supervisors. In order to mitigate these effects, dialogue within the training relationship would need to continue to be fostered (e.g. through WBA verbal dialogue as mentioned on page 125), giving supervisors information to adapt their repertoires for improvisation as the nature of trainees and practice changed. As Croskerry, Petrie, Reilly and Tait (2014) recommended, minimising dysfunctional practices and maximising functional ones was important and could include adapting to generational changes. Through a process of reflection on Judgement in Action, supervisors could reassess the main priorities of practice.

The portrayal of Schön's theory (reflection in action) as Judgement in Action helped to articulate the process of how surgical supervisors made judgements in the moment. As observed by Harris (1989), these judgements were based on situational rather than objective phenomena. Judgement in Action in surgical training demonstrated that supervisor judgements had a unique quality, drawn on personal repertoires of responses. However, as also asserted by Harris about reflection in action, although Judgement in Action could lend itself to the demonstration of judgement by experts, it should not be confined to the realm of experts. In contrast to the theory of Dreyfus and Dreyfus (2005) for example, Judgement in Action in surgical training was also of relevance for trainees. Even at the early stages of training, trainees were expected to apply their level of knowledge and experience to tasks and decisions applicable to the real world, distinguished only on the basis of how much complexity was involved and hence how much supervision was needed.

As covered in this chapter, the process of how participants tended to make judgements in the moment was called Judgement in Action. Judgement in Action had a particular pattern, recognising trainee performance in four ways (competence, decision-making, administration and relationships). This could suggest that any assessment regime which took account of this pattern of judgement might be more likely to appeal to supervisors and be more easily adopted. The next chapter explores how far Judgement in Action and CiPs complemented one another as a result of analysing the way the participants related their experience of working with trainees through the lens of three individual CiPs.

Chapter 9: Findings and Analysis
(Judgement through CiPs)

> "If programmes really move towards observing and qualifying the competence of individual candidates for critical professional activities, instead of assuming competence at the end of a predetermined training period, a paradigm shift will occur".
>
> (Ten Cate, 2005, p. 1177)

9.1 Introduction

Ten Cate's assertion that capabilities should be assessed by operationalising competencies and linking them with professional activities offered to resolve disputes about the value of competencies and to enable supervisors, trainees and the public to recognise precisely what a competent doctor could and could not do. Section 2.2 above set out how the acceptance of this concept had led to the development of CiPs in surgical training. CiPs would foster the deliberate granting of responsibilities by the gradual removal of supervision scaffolding and meant that performing well in these nine relatively few, broad, critical areas of professional practice were the mark of readiness for doctors to transform from trainees to competent surgical consultants, the ultimate aim of the surgical training programme.

The participants were asked to undertake a simulated assessment on their trainees using three of the nine CiPs. Following the same analysis process described in chapter 6, individual codes were grouped into categories and themed in answer to the research questions posed for the second stage of the study (see table 3 on page 60). This section describes how the participants

responded, how far their perspectives on their trainees altered and how far these differences conflicted with or enhanced their Judgement in Action.

In the second stage of the study, the first research question (2a) asked to what extent supervisors' perspectives of trainees altered through the use of CiPs. To answer this question, the analysis attended to any apparent variations in how supervisors described trainee performance. If any differences were identified, a consequent question needed to explore whether these added value to judgement patterns. While the content of CiPs signalled the areas of performance on which the participants could reflect e.g. providing supervisors with cues about particular tasks and skills, the participants had the latitude to be able to select which from the many cues to attend to and elaborate upon.

The addition of CiPs to the curriculum assessment system was generally well received by the participants who saw it as a means of providing useful criteria against which to judge trainee progression. This notion was similar to the use of decision aids mentioned in the literature on judgement in chapter 4. Four broad themes emerged from how participants talked about their trainees using CiPs, they were:

Theme 1: Controlling time

Theme 2: Broader training perspective

Theme 3: Professionalism

Theme 4: Junior/senior transition

Each of these themes is explained below showing links with Judgement in Action sub-themes.

Theme 1: Controlling time

The ability to control time and manage competing demands well was an often-under-rated ability that could make the difference between moving forward in the training pathway or stagnating. Trainees might for example tend to find it difficult to disengage themselves from a patient who is bleeding superficially in order to be present at theatre to observe a rare procedure. Here, a moral question of '*what ought I do?'* and the judgement exercised by trainees echoed Coles and Fish's deliberative judgement (1998). In surgery there was an added dimension about choosing which actions *not* to take. Trainees who could not be dispassionate in this way might never expect to reach consultant grade. This theme corresponded with Judgement in Action and decision making in particular. Two related categories under this theme were *trainee self-management* and *trainee understanding of timing for intervention.*

TRAINEE SELF-MANAGEMENT: The current surgical curriculum emphasised the role of trainees in managing their learning, CiPs correspondingly included competencies about managing time e.g. *CiP 1: Manages time within the clinic setting*. Participants 4 and 5 reflected specifically on how far their trainees were able to balance the demands of different environments e.g. to plan time to see all the patients on the ward and depart in time for the start of a theatre or clinic. As shown in the literature on trust in sections 4.5 and 4.6, organisational skills, including organisation of oneself, were often seen as a proxy for trustworthiness (Hauer *et al*. 2015). Correspondingly, Participant 4 saw one trainee who had

blamed the system for her failure to attend events as less reliable than another who was perceived to have managed to overcome such obstacles.

Participants 4 and 5 drew attention to trainees recognising the importance of clinics and making time to attend them. Currently trainees tended to want to spend most of their time in theatre because it was felt to be exciting. However, clinics provided vital background information about patients and their needs. Participant 4 laid emphasis on successful clinical outcomes “all coming down to patient selection and management of clinic beforehand”. Using Camerer and Johnson’s argument (1997), set out in chapter 4, expert insight could be created through situational knowledge of rare factors (also known as broken-leg cues) e.g. about individual patient issues, needs and preferences that could impact on their treatment and recovery. Participant 4’s insight echoed the view of participant 6 whose original interview expressed the risks to patient safety from trainee absences from clinics. The emphasis on the importance of making time for learning in different environments, provided by CiPs, appeared to add value and could also help to ensure trainees achieved sufficient training in these areas in future.

All the participants, with the exception of participant 2, mentioned that CiPs highlighted the importance of trainees managing their time when working in each training environment. CiPs highlighted the need for time efficiencies and prompted participant 1 to judge the junior trainee as better able than the senior trainee to manage time in clinic because the senior trainee spent too much time on detail and was “too careful”. CiPs, therefore, also appeared to differentiate

abilities that were unrelated to the training level and could mean that supervisors were able to take an ipsative view to scaffolding trainee learning (Hughes, 2014).

> "'Managing time in the clinical setting' - I would put [the more junior and senior trainee] the other way around. [The more junior trainee] is very efficient in managing clinic time. [The senior trainee] is very detailed, sometimes more detailed than what is needed". (P1)

TRAINEE UNDERSTANDING OF TIMING FOR INTERVENTION: Participants 1 and 4 felt that junior trainees tended to employ too many tests on each patient. They also highlighted that all but the most senior trainees tended to order investigations too soon for conclusive evidence of pathology. Consultants and advanced trainees had the confidence to wait for symptoms to appear and make better use of scarce resources.

> "[The more junior trainees] may do a CT [scan] more quickly than a consultant who will have said wait for that [symptom to appear] and see if that's alright". (P1)

> "… because of the rarity of the condition they [trainees] would just not know what they were looking at and that patient would probably end up having to come again or they might end up with unnecessary investigations cos they don't recognise what they are looking at." (P2)

Referring to the model put forward by the Dreyfus brothers (2005) outlined in chapter 4, trainees acting as the *competent performers* or *proficient performers* followed rules to manage the information overload while experts, in contrast, were more realistic in their actions, had an extensive repertoire of situational experience on which to draw and would also seek alternative views by consulting others. Arguments in the literature that predictive accuracy was an appropriate measure of expert judgement (Camerer & Johnson, 1997) appeared to be less

important to the participants than critical thinking and problem-solving (Coles, 2002)

Participant 4 stressed the need for clinicians to understand the differences between emergency and routine cases and the importance of dovetailing test results and clinic appointments to avoid unnecessary delay in treatments. This necessitated knowledge of signs and symptoms, how long tests were inclined to take, the likely treatment path and the ability to compel urgent treatment for serious cases. This was an area in which participant 4 felt that trainees' knowledge tended to be deficient.

> "knowing an incisional hernia that's sore isn't an emergency, and can wait for a routine appointment, knowing that weight loss anaemia is a red flag and needs to be seen urgently or if they've got … indigestion, heart burn and weight loss that the most efficient way to get a result is to book an endoscopy and a clinic appointment at the same time, so that you can get the scope and see them with the result … so knowing that they [trainees] understand that and the pathways, … I think they wouldn't know if I said, 'How long does it take for an endoscopy?' 'How long for a CT?'." (P4)

Participant 5 highlighted the importance of the clinician's ability to progress-chase theatre support staff (corresponding with administration and relationship capabilities). If trainees left the theatre in between patients, the support team would see their absences as reason to delay patient preparation. Trainees needed to be aware that they had the responsibility to actively supervise patients' throughput.

> "Also keeping the list moving involves you being there. I often do say to the trainees that if when one operation finishes, and you disappear and leave the operating theatre and wait for them to call you when the next patient's on the table, you're going to lose a few minutes in

between every case because you don't keep saying, 'Have we asked for the next patient? Have we asked for the next patient? Is the next patient on their way?" (P5)

Theme	Categories	Code	Link with Judgement in Action
Controlling time	Trainee self-management	Trainee ability to manage themselves and their time in order to obtain training opportunities. (P4 / P5)	Decision-making
		Trainee ability to manage time in theatre, ward and clinic (P1 / P3 / P4 / P5 / P6).	Decision-making
		Trainee understanding of their role with that of others in streamlining the whole patient journey. (P4)	Relationships
	Trainee understanding of timing for intervention	Junior trainee tendencies to over-test. (P1 / P4)	Competence Decision-making
		Trainee ability to gauge the right time to run tests. (P1 / P4)	Competence Decision-making
		Trainee understanding of the nature of emergencies and the need to actively ensure treatment. (P4)	Relationships Decision-making Administration Competence

Table 9: Theme 1 - Controlling Time

Summary of theme 1

Participants' judgements through CiPs appeared to prompt a deeper level of thinking about the necessity for trainees to control their own and others' time. Controlling time meant more than the ability to manage time as it entailed the self-discipline and judgement of trainees whether to actively pursue and persist or

remain patient in difficult situations. This theme also resonated with the notion of Judgement-in-Action in terms of making decisions about one's own time and that of others, about when to intervene in the patient journey through an understanding of how individual roles related to other elements of the treatment pathway.

Theme 2: Broader training perspective

While surgical training taught specialised knowledge and skills, clinicians were expected to have a peripheral view of the context within which the treatment they offered was grounded. Consequently, the CiPs initiative encouraged supervisors to take both a longer and more integrated view of trainee abilities so that they would be judging holistic activity and encouraging trainees to have a wider view of the context in which they worked. The participants agreed that there was benefit in viewing the ability of trainees to undertake broader activities that were made up of multiple tasks. Under this theme, they commented on how they tended to judge holistic activity and what the wider view might entail.

JUDGING HOLISTIC ACTIVITY: Participant 2 viewed trainee responsibilities as far more difficult in the light of CiPs in which they were assessed on managing a chain of events incorporating assessing the patient, preparing them for theatre, performing the operation and post-operative care.

> "Our trainees have to assess the patient, take them to theatre and do the operations while they are still managing all that's going on and it is a difficult job and they are rung up constantly." (P2)

CiPs were designed to enable trainees to be viewed against the end-point of training while all the participants followed the current system of viewing trainees

by stage of training. Participants 2 and 4 commented that CiPs helped them to see trainees in light of their readiness for certification. Their focus changed to asking themselves whether they would appoint the trainee in question to a consultant post and if not, why not. However, consistent with their use of Judgement in Action, they considered that poor relationship skills would veto technical competence when making decisions about appointments.

> "There is a whole list of things you do sometimes consciously and sometimes sub-consciously. The one about end of training was if we had a job would we appoint this person … Sometimes there are people who are perfectly good, but you wouldn't appoint because they are awkward characters or something but usually that thought 'would we appoint this person?' Which also goes with it … 'would I be happy to leave my patients with this person looking after them?'." (P2)

CiPs led participant 4 to reflect on the need for trainees to be able to manage the 'remote' patient, an area of responsibility not previously mentioned. Some patients never attended hospital but were treated by GPs under the advice of the hospital specialist. Through telephone conversations that dealt with the GP's questions, the specialist could explore the patient history, advise on tests and help to analyse results. The clinician could build a relationship with the GP and be involved with ongoing issues without ever seeing the patient. Participant 4 felt that this was an important task to which trainees tended not to be exposed in the current system.

> "You might give them some advice to try some medication or it's simply a hangover from what they had before, there's nothing different to do. The patient never comes near the hospital, so the trainees would never see those patients and wouldn't be involved in those discussions. If it was a bad problem that they came in as an emergency, then they'd see those patients but that's the tip of the iceberg ..." (P4)

WIDER VIEW: A competency in CiP 2 quoted below, led participants 1 and 4 to raise important differences in trainee ability to manage palliative care.

> "Recognises when further therapeutic manoeuvres are not in the patient's best interests, initiates palliative care, refers for specialist advice as required and discusses plans with the patient and their family."

In the workplace, surgical judgements about the terminally ill carried with them a moral dimension, reflecting the phronesis (practical wisdom) type of judgement asserted by Coles on the nature of professional judgements (2002) and again the moral question '*what ought I do*?' suggested by Coles and Fish as relevant to deliberative judgement (1998). Junior trainees tended to think in terms of percentage rates of survival after an operation, befitting the risk society viewpoint of Beck (1992) alluded to in chapter 4. However, according to the participants, it was also important to consider post-operative quality of life. Junior trainees were more likely to persevere with treatment while senior trainees were more confident about judging, in concert with the patient and relatives, when and how to move to palliative care. This was an aspect of clinical decision making that strongly corresponded with the decision-making sub-theme of Judgement in Action (see table 10).

> "When you are a junior registrar you think yes there is a 5% or 10% chance of survival, there's no chance of survival … but that is not appropriate all the time because the reason is you can operate on the patient and the patient survive but the quality of life will not be good and … It is not just about the resources it is about the quality of care." (P1)

Participant 6 felt that trainees tended to be sheltered from and deficient in their knowledge of how to navigate the political landscape within which NHS hospitals existed. While in training, doctors tended to be protected from global initiatives

such as waiting times and implications of breaches. It was only when trainees became consultants that they were expected to work within the parameters and manage the side-effects of governmental initiatives (such as over-booked clinics). Participant 6 felt that trainees needed to develop an understanding brought about by exposure to these realities. In addition, participant 6 felt that trainees tended to be deficient in their knowledge of legal aspects of care and information governance. Although senior trainees were expected to develop this knowledge through required reading, they tended to be weak in this area.

> "So, the one thing the trainees don't get, and I never had either is how the wider aspect of the NHS and how it works. We don't do that, as a trainee you concentrate on your individual patients and your firm, and maybe the ward and that's it … I don't think any trainee has it." (P6)

Theme	Categories	Code	Link with Judgement in Action
Broader training perspective	Judging holistic activity	Trainee ability to cope with a chain of events. (P4)	Competence Administration Decision-making Relationships
		Viewing trainee as appointable consultants. (P2 / P4)	Competence Administration Decision-making Relationships
		Trainee ability to manage the remote patient (care in the community). (P4)	Competence Administration Decision-making Relationships
	Wider view	Differences in trainee ability to manage palliative care. (P1 / P4)	Decision-making

		The need for trainees to be able to manage the political landscape. (P6)	Relationships Decision-making
		Trainee knowledge of legal aspects of care and information governance. (P6)	Competence Decision-making

Table 10: Theme 2 - Broader View of Trainees

Summary of theme 2

Participants' judgements through CiPs appeared to help them appreciate trainee performance in the context of joined-up activities and 'appoint-ability' as consultants. Trainees needed to be introduced to rather than protected from the political, economic and legal realities of the system in which they worked. This theme mapped well to all four elements of Judgement in Action (see table 10).

Theme 3: Professionalism

When participants used Judgement in Action, they rated professional skills, such as the ability to communicate and form good working relationships, very highly. Good social skills were suggested by Shanteau as one of the characteristic features of experts (1988) as discussed in chapter 4. However, a previous study (Bussey, 2017), established that supervisors tended to be less confident when assessing professional skills such as the ability of trainees to deal with and reflect on complaints, challenging patients and unforeseen problems. Participants felt that trainees needed to take personal responsibility for their work such as by verifying what they did and preparing in advance. This theme corresponded well with the competence, decision-making and relationship sub-theme of Judgement in Action as shown in table 11.

THE GENERIC PROFESSIONAL CAPABILITIES (GPCs): The GPCs incorporated within CiPs were intended to encourage supervisors to explicitly assess professional skills by giving them the vocabulary (descriptors) they needed.

Participant 4 recognised that clinical and professional skills (while entwined in daily activities) were independent of each other and that junior trainees could for example have good professional skills (such as the ability to listen and explain) without the requisite clinical knowledge. As discussed in chapter 4, Ericsson and Smith (1991) had hypothesised that stable personal characteristics could be either inherited (e.g. personality) or acquired (e.g. skills) and Shanteau (1988) also suggested that experts had certain positive personal characteristics. While expertise tended to be acquired, participant 4's observation below might suggest that certain personal characteristics could speed up or slow the attainment of professional skills.

> "… the communication skills … a fairly junior doctor who's very good at, you know, doing all those right things but just haven't yet got all that clinical knowledge … they speak to the staff, they speak to the patients, they explain things well." (P4)

Participant 1 felt that the ability to reflect on one's practice was an important element of dealing with complaints, as well the ability to apologise.

> "… you can head things off at the beginning by apologising and accepting responsibility and actually the drain into the bowel has led to a complaint … and part of the problem at the beginning is that [the trainee] didn't apologise for it." (P1)

Participants 1, 4, and 5 commented on 'emotional resilience', a descriptor in GPC domain 1 (appendix 7). They had varying views about the meaning of this ability,

feeling that more clarity was needed before it could be assessed. Participants 1 and 4 felt that reflection on complaints could be seen as a sign of emotional resilience because it could evidence dispassionate review of difficult circumstances and the ability to make changes to improve practice, suggestive of deliberation for the intention of improvement outlined in chapter 4 (Litzinger, Lattuca, Hadgraft & Newstetter, 2011). Participant 5 felt that emotional resilience had meaning in relation to dealing with challenging patients, those who might be particularly intelligent and more demanding, the young and those that trainees could particularly relate to, such as a pregnant woman with cancer. There was an element of learning how to distance oneself, endorsed by participant 5 who felt that dealing with events that did not go to plan because of complications needed to involve both reflection and recovery.

> "… it's how they respond to criticism and when things go wrong I think". (P4)

> "You do have to learn to distance yourself from it a bit as well. Again, another sort of resilience you need in surgery is being able to deal with things that don't go according to plan – with complications because everybody has complications … It's obviously depressing for the patient and it's demoralising for you as the surgeon … [we have to look] back and reflect and see if we could learn from it". (P5)

PERSONAL RESPONSIBILITY: Participant 5 referred to trainee ability to complete clinical documentation (mentioned in CiP 1) and the importance of trainee sense-checking or overviewing clinical reports when making a diagnosis or treatment plan. Trainees needed to refer to source material (such as a pathology report or x-ray) rather than relying on secondary information such as a letter from a GP. As discussed in chapter 4, the literature on judgement suggested that decisions by experts involved systematic review of information from rigorous data-

gathering. In a similar vein, participant 5 felt that patients should be seen as a primary source of clinical information. Often trainees did not take the time to speak to the patient before their operation. In participant 5's view, training would be more effective if trainees did more to prepare in advance for planned procedures.

> "I've had a few letters come back from GPs saying this doesn't make sense … If they haven't checked on information from previous encounters, then they might have got the wrong information about a previous cancer or something like that and the GP's questioned it. I usually tell people … to go back to the primary source." (P5)

Theme	Categories	Code	Link with Judgement in Action
Professionalism	Generic professional capabilities	Recognition of trainees' professional skills as an independent set of skills. (P4)	Competence Decision-making Administration Relationships
		Trainee ability to reflect on complaints (P1 / P4).	Relationships Relationships (with self)
		Trainee ability to deal with challenging patients (P5)	Competence Relationships
		Trainee ability to deal with unforeseen problems. (P5)	Competence Decision-making Relationships (with self)
		Trainee ability to reflect on complications. (P5)	Competence Relationships (with self)
	Personal responsibility	The ability of trainees to verify the accuracy of their work. (P5)	Decision-making

		Trainee disposition to prepare in advance. (P5)	Decision-making

Table 11: Theme 3 - Professionalism

Summary of theme 3

Overall, CiPs appeared to promote the assessment of professionalism. Through the GPC element, participants considered aspects of trainee professionalism not previously mentioned, they included the ability to reflect on difficult situations and take personal responsibility. Trainees needed to appreciate the gravity of serious events but also be able to consider how to improve, they needed to be able to recover in order to keep practising. To help ensure professionalism, trainees needed to take personal responsibility which included sense-checking using primary sources of information, being patient-centred and preparing in advance. These comments corresponded with the competence, relationship and decision-making elements of Judgement in Action as shown in table 11.

Theme 4: Junior/senior transition

The CiPs prompted the participants to recognise some further differences between junior and senior trainees that had implications for how they should be trained and judged. This theme corresponded well with the competence and decision-making sub-themes (both core to surgery) and to relationships (complementary to surgery) of Judgement in Action. Trainees had to be able to adapt to complexity by understanding the nature and signs of emergencies while planning for complications, they had to be able to communicate with a range of colleagues, understanding what different team members needed from them.

Trainees in different stages of training had different responsibilities and their remit had implications for how they performed in different training situations.

ADAPTING TO COMPLEXITY: This theme echoed Schön's characterisation of professional work as uncertain and unpredictable. It also resonated with Shanteau's (1988) distinction between substantive and assessment experts. Clinicians' skills were of the latter because they lay in making judgements despite uncertainty and required perceptual, diagnostic and action expertise. Participant 4 felt that trainees had to take responsibility for planning ahead for complications that might arise (such as operations not proceeding in the expected way). Trainees needed to be judged on whether they could draw on alternative strategies (closely corresponding with the notion of improvisation mentioned in chapter 8, page 159). Trainees could only obtain these skills with experience of difficult cases which tended to occur rarely.

> "We know in our heads the operation's going to go this way and if it doesn't we have strategies for dealing with that, but how we assess the trainee's ability to adapt, cos often you come in with plan A and if it changes to plan B it becomes your plan B you'd have to consciously work out whether the trainee had a plan B." (P4)

COMMUNICATION: Participant 4 felt that junior trainees, more senior trainees and very senior trainees tended to communicate with supervisors differently. Junior trainees needed to check every detail with their supervisors while more senior trainees tended to know and do what needed to be done and would feel inclined to merely reassure supervisors that everything was in hand, often to 'prove they could act up' (P4). Very senior trainees, however, were likely to be more communicative with supervisors not because they needed help but out of the courtesy of keeping senior colleagues up to date.

> "The fellows [very senior trainees] who'd seen more and were kind of fledged but just needed a little more experience, knew all the things that you would want to know, and would come and tell you, 'This doesn't look quite right…' the others [less senior trainees would say] 'It's OK it's all in hand the patients are roughly fine.'. They'll keep the details to themselves more than somebody who's further on ..." (P4)

This finding highlighted the need for good communication in the trainee-supervisor relationship at each training level. For example, when trainees started a placement in a new unit, the first meeting between trainee and supervisor could include a discussion about preferred ways of communicating.

Participant 6 felt that trainees needed to develop an in-depth understanding of how to communicate with associated specialists, an ability that came with time and experience, corresponding with relationships sub-theme of Judgement in Action. Surgeons and anaesthetists, for example, worked closely and had to appreciate the implications of knowing the other's methods and techniques and the timings involved. Participant 6 felt that trainees tended to be unaware of the nuances of the consultant and anaesthetists "casual talk" in theatre. Consequently, when trainees were judged ready to lead the theatre team, their ability equated to more than confidence and competence in doing the procedure and the ability to use assistants well, they had to have learned the critical communication involved in undertaking advanced procedures.

> "… for [the junior trainee] with an EVT [operative procedure], even though he's very junior he'd be perfectly capable of taking that leadership role, for [procedure] [but] he had no idea that my casual, 'The tumours out,' … is actually an indication to the anaesthetist [to lower the blood pressure]." (P6)

JUNIOR / SENIOR REMIT: Participant 6 made a useful distinction between the knowledge and skills of junior and senior trainees owing to their stage of training. Junior trainees might be viewed as better than seniors at managing a ward round while senior trainees were better in clinic and theatre. Junior trainees were mainly situated on the ward and tended to know all the patients well but could not manage all the patient problems. More senior trainees, like consultants, knew less about what happened on the ward but more about individual patients needs and how to manage their treatment. Different criteria, therefore, needed to be borne in mind when judging trainees at different levels of training.

> "I think our junior trainees are perfectly competent doing a ward round, like a senior trainee, except they don't have the individual knowledge of the condition enough to say… so, they'd know all the patients, they'll know the diagnosis and they'll have a plan but [the junior trainees] don't know that a [surgical procedure] needs morphine for 21 days and then it should be stopped or they haven't read the research that shows that if you haven't had a seizure then [drug name] should only be given for 1 week and they won't know to stop it." (P6)

Theme	**Categories**	**Code**	**Link with Judgement in Action**
Junior/senior transition	Adapting to complexity	Trainee ability to plan ahead for complications (P4).	Competence Decision-making
		Trainee understanding of emergencies, the signs and approaches. (P4)	Competence
	Communication	The difference in approach between juniors and seniors. (P4)	Relationships

		Trainee understanding of how their supervisor wishes to communicate. (P4)	Relationships
		In-depth understanding of the roles and implications of communication with associated specialists. (P6)	Competence Relationships
		Differences in leadership in theatre. (P6)	Decision-making Relationships
	Junior / Senior remit	Differences in trainee responsibilities for ward, clinic and theatre at each stage of training. (P6)	Competence

Table 12: Theme 4 – Junior/Senior Transition

Summary of theme 4

CiPs appeared to promote the assessment of planning for uncertainty, communication with seniors and colleagues and the understanding of different remits along the training pathway. Participant comments corresponded with the competence, relationships and decision-making elements of Judgement in Action as shown in table 12.

Overall summary

Evidence from the interviews suggested that CiPs were successful in prompting new ways of thinking about trainee development and that these fell into the four components of Judgement in Action. Participants' reflection on trainee judgement changed subtly from the pre-CiP analysis in four broad areas. These areas revealed that participants perceived as well-performing trainees those that could control time (which was more than just managing time), had a broader perspective of the health system, were prepared for challenging situations and

showed they could transition from junior to senior responsibilities. Tables 9 to 12 summarise where the findings from each of the themes correspond with the four areas of Judgement in Action, for example, the need to control time in table 9 was an essential element of all four capability categories of competence, decision-making, administration and relationships. Most of the insights prompted by CiPs centred on the relationships and decision-making categories. These were the higher order elements lending support for the view that assessing trainees against the end-point of training to gauge readiness for consultant practice was important and hence helping to validate the rationale behind the CiPs (outcomes-based) approach.

The second research question (2b) asked how far the CiP approach resembled Judgement in Action. The CiP approach was defined in section 2.2 as a summative, holistic assessment that aimed to draw on the subjective professional view of the supervisors with whom trainees worked. These professional opinions were based on knowledge of both the trainees and of clinical practice. The CiP approach was outcomes-based, focused on the end-point of training defined as readiness for independent consultant practice, and realised through the lessening need for supervision. Judgement in Action, however, was shown in this study to be what supervisors actually tended to do but in an unconscious and unsystematic way. If CiPs was to work successfully, it would need to enhance the use of Judgement in Action and not work against it. Comparing the use of CiPs with the participants' use of Judgement in Action indicated six ways in which CiPs converged with Judgement in Action and one way in which it diverged as outlined below.

Areas in which Judgement in Action corresponded with CiPs included:

TRAINING RELATIONSHIP: Both approaches relied heavily on the existence of a training relationship between trainee and supervisor. All the participants described clinical situations where they worked together, gaining knowledge of each other's styles and techniques (as shown under *trainee-supervisor relationship* in the coding table in appendix 8).

LEVEL OF SUPERVISION NEEDED: All the participants felt that that trainee performance could be differentiated by how much help trainees needed from senior colleagues, corresponding with the notion of a decreasing level of supervision (as shown under *Judgement in action, nature of supervision* in the coding table in appendix 8).

KEY ENVIRONMENTS OF WARD, CLINIC (INCLUDING A&E) AND THEATRE: All those taking part in the study described training situations in terms of the three main environments of ward, clinic and theatre(as shown under *environment* in the coding table in appendix 8). Participants highlighted the centrality of the management plan in the treatment of patients, strongly featured in CiPs.

HOLISTIC VIEW OF TRAINEES: The participants described practice in terms of complexity and difficulty and awareness of context (as shown under *judgement in action* in the coding table in appendix 8).

SUBJECTIVE JUDGEMENT: (although currently implicit)**:** Participant judgements tended to be based on opinions (both their own and those of their peers/colleagues) that they had developed through the experience of working

with trainees based on their sense of what it meant to be a surgeon (as shown under *trainee-supervisor relationship* in the coding table in appendix 8).

REFLECTION ON PROFESSIONAL ABILITIES: Supervisor judgements involved examples of good professional skills such as reflection on serious events, the need to recognise the limits of one's competence, the need for insight, commitment, openness to feedback, accuracy, inquisitiveness, self-management, initiative, leadership, confidence, good decision-making and keeping knowledge up to date (as shown under *trainee qualities* in the coding table in appendix 8). They also highlighted the importance of relationships and multi-disciplinary team-working (as shown under *sources of evidence* in the coding table in appendix 8).

Area in which Judgement in Action did not correspond with CiPs:

JUDGEMENT IN ACTION WAS INFORMAL: CiPs would formalise supervisor judgements through the new structured CiPs assessment with the attendant risk that supervisors might be reluctant to provide formal judgements (such as a supervision level) when objective evidence was thin. The participants felt that it might be difficult to ever justify awarding a level 4 when there was a possibility that trainees might go on to perform poorly.

> "… so, at the ARCP … they're going to say, 'Show me the evidence, where's your evidence', and there isn't any, if there isn't any it's just going to get overturned …" (P6)

Analysis conclusion

Overall the CiP approach appeared to correspond well with participant Judgement in Action. CiPs reflected the supervisor-trainee relationship that

involved the lessening of supervision as proficiency increased. CiPs also corresponded with the key tasks, activities and environments being judged as well as the type of professional qualities supervisors looked for in trainees. There was one area in which CiPs corresponded less well with Judgement in Action. This was related to the formalisation of the judgement process which was at odds with the informal, subjective manner in which participants tended to make judgements. While not posing a serious conflict, it was an issue that would need to be addressed through training if CiPs were to be introduced.

Chapter 10: Contribution to theory and practice

The following section outlines the contribution Judgement in Action makes to the theory of judgement and to the practice of assessment. It also sets out three options as recommendations for the design of CiPs. Finally, it outlines the study limitations and summarises the anticipated next steps towards both the further exploration of Judgement in Action and the introduction of CiPs.

10.1 Introduction

A test of merit of the theory of Judgement in Action would be in its ability to be able to make a contribution to both theory and practice. Areas in which this might apply are outlined below.

Theory

Building on Schön's theory of reflection in action (1990), this study established that there was a common pattern of Judgement in Action across the six participating supervisors. Judgement in Action was subjective, informal and based on collaborative working, allowing unplanned but intentional learning to occur in the workplace in response to imminent situations as they unfolded. Judgement in Action referred to a holistic perspective of trainee performance that combined competence, decision-making, administration and relationships. It involved the supervisor having a concrete experience with a trainee which generated new understandings upon which to consider prior understandings about the trainee's ability. This in turn led to experimentation about the degree of supervision required and, therefore, the amount of trust granted to the trainee in

turn leading to further experiences and understandings. Notwithstanding other environmental factors influencing supervisor judgements, in line with Kolb's theory of experiential learning (1984), Judgement in Action was an active process.

The theory of Judgement in Action offers a great advantage over the current position in which supervisors tend to be unaware of the nature of their own judgement practices and where there is no guarantee that assessment methods usefully draw on them and, therefore, truly engage supervisors in meeting the objectives of assessment. As a result, Judgement in Action has the potential to inform multiple stakeholders. Supervisors might find it useful to reflect on the theory, their possible unconscious tendencies and whether their personal judgement pattern balanced each of these parameters in appropriate measure in order to improve their practice. Trainees, as learners, may have particular interest in knowing how they tended to be judged and in what elements they might need to develop proficiency in order to be seen as progressing. Finally, curriculum developers may find the theory useful when developing WBA (see section 10.2 below). Moreover, the theory may be of interest to a wider audience than those involved in surgical training.

Although the theory emerged in surgery where fast decision-making under pressure using limited information in a multi-disciplinary environment was needed, it could be applied to training in other fields. The extension of the theory would mean that while the four overarching capabilities remained unchanged, the underlying features of those capabilities could be unique. For example, a criminal investigator working for the police service would need to be trained to follow legal

protocols (competence), make considered judgements based on documented evidence and prioritise lines of enquiry (decision-making), work with professionals in allied fields (relationships) and manage a complex investigation (administration). Alternatively, in airline industry training, recruits would require technical expertise and knowledge of protocols (competence), the ability to communicate with and co-ordinate ground and flight teams (relationships), ability to follow a documented process (administration) and make rapid decisions to cope with system and human errors as well as security threats (decision-making). This theoretical pattern of Judgement in Action might, therefore, serve well as a practical approach to the judgement of learner performance in a variety of professions and could be explored further.

Practice

An aim of the study was to explore whether CiPs drew more closely on supervisor expert judgement and in so doing, could restore a more holistic view of surgical capabilities leading to a more complete picture of trainee performance. This study showed areas where CiPs and Judgement in Action did and did not correspond (see pages 184-185). They corresponded through the provision of trust through the lessening of supervision as proficiency increased, the reflection of the key tasks, activities and environments being judged as well as the type of professional qualities participants looked for in trainees. By reflecting Judgement in Action, CiPs might help to enhance judgement by making it more explicit and systematic, leading to a greater degree of understanding and trust between trainees and trainers and effort being directed at what was important. By showing that CiPs incorporated the critical components of Judgement in Action, the GMC as regulator might be satisfied that curriculum change met its standards (2017).

Taking account of Judgement in Action in the design of any assessment regime might also suggest an appreciation that an authentic pattern of judgement might be more likely to appeal to and be adopted by supervisors. Supervisors and trainees who might see assessment in the workplace as difficult, irrelevant or a threat to their self-concept may soften their attitude when the practical application of the theory was recognised as an innovation that was responsive to their needs.

A second practical contribution lies in the identification of flaws in CiPs and the suggestions for tackling them (see section 10.2 below). While Norris (2014) proposed that curricula could be incrementally improved, in practice, once implemented, further change can be disruptive. The final practical contribution consists of framing Judgement in Action as a driver for a change of views about assessment. The implications for assessment include recognition of the importance of human processes, awareness of the judgement pattern in use and a considered approach to its integration into assessment. The practical application of Judgement in Action might be of interest to curriculum developers in other medical and non-medical fields facing the same mandate for curriculum revision in postgraduate medical education.

10.2 Recommendations and next steps

The study aimed to assist the move to an outcomes-based curriculum which re-established the professional opinions of supervisors in workplace assessment. Assessment systems aim to ensure assessors make justifiable decisions about learners. Those who apply them use human processes and are more likely to engage if the systems themselves draw on the judgement patterns they use. The

emergence of Judgement in Action created an opportunity to integrate human judgement with the new CiPs assessment and three options are outlined below.

OPTION 1: The study showed that Judgement in Action corresponded well with CiPs by drawing on the training relationship in which trainee and supervisor worked together. Help for trainees was scaffolded according to the uncertainty and complexity of different workplace situations. Judgements were holistic and subjective, and the qualities judged were both clinical and professional. However, small modifications could be made to bring about an even closer match. The four elements of Judgement in Action; competence, decision-making, administration and relationships could be reflected more strongly in the content of each specific CiP. For example, CiP 4, *Managing the Operating List* (appendix 6), mentions the need to interact appropriately with members of the team and to lead the de-brief. However, the need to build good relationships and an understanding of the roles of others, especially between the surgeon and anaesthetist, could be further emphasised. In addition, CiPs might more closely reflect Judgement in Action if the assessment was formative.

OPTION 2: An alternative to strengthening CiP content by incorporating elements of Judgement in Action would be to re-structure CiPs into a quadrant format, reflecting each of the four Judgement in Action capabilities. Each of the nine CiPs could be formatted in this way. This format might help supervisors make connections between the assessment and their unconscious judgement tendencies so that CiPs appear more user-friendly.

OPTION 3: Going a step further, the nine re-formatted CiPs could be replaced with one generic CiP, covering all training situations. Under each capability there could be a set of generic competencies to fit any training situation. However, breaking away even further from the notion of CiPs as decision aids, a new approach could be adopted which would allow supervisors unrestricted judgements by removing the prompts under each capability, leaving simply an open space for free-text.

Further research could be undertaken to explore the practical application of Judgement in Action to CiPs. Studies could also be carried out to explore how the theory of Judgement in Action might apply to the way supervisors in other professions tend to judge learners. While the evaluation of the practical implementation of CiPs fell outside the remit of this study, it nevertheless aimed to inform the feasibility of this new assessment tool. Trialling CiPs in the workplace (including the utility of supervision levels), using for example, Kirkpatrick's (1979) four levels of: reaction, learning, behaviour and results, would be necessary to further enhance the educational validity of CiPs.

10.3 Limitations

The nature of the study gave rise to a number of limiting factors which need to be acknowledged when arriving at any conclusion based on the study outcomes.

INSIDER/OUTSIDER POSITION AND CONFLICT OF INTEREST:

a) The study was undertaken from the viewpoint of curriculum development (researcher as an 'outsider' in parallel with Schön's 'high-ground') rather than from within the profession (researcher as an 'insider' in parallel with Schön's 'swampy lowlands') and, therefore, the researcher might have had an

unconscious bias towards the logic behind the design of CiPs. This perception might have led participants to be unwilling to share the messy realities of clinical practice. However, the researcher's knowledge of the research setting and understanding of its wider context allowed the researcher to understand the sensitivities involved. Participant accounts were central to the study findings and analysis.

b) The research followed principles and processes laid down by UCL ethical standards which formed the fabric of the participant-researcher relationship. The researcher had no direct dealings with the participants or influence on their status, minimising any implication on the participants' practice. Tensions were also eased by the endorsement of the Royal Colleges of Surgeons, which lent credibility to the aims of the study and without which the endeavour would not have been possible.

THERE WAS NO OBSERVATION OF PRACTICE: The study design did not include any observation of surgical training as a data collection strategy. The need for patient consent on a large number of occasions and the inability to control what might be observed made this impractical. This was considered as not weakening the authenticity of the study, however, as it was considered likely that supervisors made summative progression decisions privately, rather than overtly, within the clinical workplace. Therefore, interviews were seen as an appropriate instrument for raising judgement decisions to consciousness.

THE STUDY WAS NOT LONGITUDINAL: The study did not follow the judgement-making process as it evolved. However, supervisors were interviewed at a real critical progression waypoint when they were thought to be ready to make these decisions about real trainees. The study asked supervisors to draw on specific examples of performance about the trainees under their care to illustrate their thoughts.

INTERVIEWS WERE THE MAIN SOURCE OF EVIDENCE: The study relied only on interview data. However, these were in-depth and in two parts and employed a simulation exercise to help validate participant perspectives from two different perspectives.

THE MOCK ASSESSMENT USED THREE OF THE NINE CIPS: Participants undertook a theoretical experiment. These represented three primary areas of patient care (clinic, ward and theatre). Although the assessment was authentic in format, it may not have revealed as many of the complex operational implications as an official assessment. Future pilots and longitudinal studies of CiPs in use will, therefore, be needed.

THE PARTICIPATING GROUP WAS SMALL: Participants were surgical consultant supervisors from a mix of specialties and different hospitals and regions with established training programmes. However, they were a self-selected group. Supervisors from other specialties and institutions might have responded differently.

Chapter 11: Conclusion

The debate about the direction of postgraduate medical education, born out of frustration with the competency approach, led to regulatory change mandating outcomes-based curricula (GMC, 2017), driving the development of CiPs. In practical terms this initiative sought to enable consultants who acted as supervisors to use their professional judgements in the assessment of trainees. No longer were lists of competencies to be relied upon to drive assessment, instead a more complete picture of performance would be required through the recognition of capabilities. Capabilities were seen as vital in a move towards helping learners make sense of uncertain situations, by drawing on sources of information, prioritising issues and arriving at practical solutions.

Schön's (1990) metaphor of the hard and swampy grounds aptly described the dichotomy of perspectives between curriculum strategy and training reality and the challenge of re-orientating assessment to balance these competing outlooks. The principles of the curriculum, arising from logic and order were most easily identifiable on the hard, high ground. However, ironically, the professional standards of the curriculum were nowhere more essential than in the swampy reality of practice, entangled as it was with compromises needed to treat patients while dealing with uncertain outcomes. The human factors that arose in situations like these influenced people and their behaviour and had a significant impact on performance and clinical outcomes (Reason, 1995). They had implications for patient safety as well as trainee and trainer well-being. In addition, in medical practice the importance of judgement and professionalism revealed themselves as particularly relevant in the critical incident, where poor decision-making could

have life-threatening implications for patients. These were ‘often highly stressful – even sometimes traumatic’ situations (Cunningham, 2008, p.161). CiPs aimed to capture from training the kind of tacit knowing-in-practice that developed through experience and through human processes like judgement. These were mechanisms that surgical professionals created in order to manage their environment but about which they ‘knew more than they could say’ (Schön, 2017, p. viii). A key principle upon which CiPs was based was the importance of placing the training relationship at the heart of the process. Supervisors made judgement decisions about trainee performance every day and were, therefore, in a position to reliably judge trainees on their ability to accomplish professional activities. The notion of an outcomes-based curriculum meant that these activities were grounded in daily practice, but now set at the standard of the end-point of training. In other words, trainees would be judged on what they would need to be able to do as a ‘day-one consultant’.

While supervisors had a global sense of whether they could leave trainees to do these activities on their own, CiPs required these to be more formally enacted through entrustment decisions translated into five supervision levels (page 53), determining the degree of oversight trainees needed to ensure they performed safely and effectively. The notion of trust, therefore, related to supervisor judgement because supervisor decisions about the level of supervision trainees required were based on how far they trusted trainees to be able to act independently. In this way, there was cognisance of the importance of balance between ‘knowing when to trust, when to trust trust and when to demand an audited account’ (Power, 1999, p.146).

This re-orientation saw both high ground and lowland areas as necessary and as needing to complement one another. The high ground was essential for drawing together examples of good practice in such areas as judgement of trainees to help unify practice and safeguard regulatory standards. However, it was from the human processes used by supervisors in the swampy lowland that good practice had to be drawn. In order to be successful, CiPs would need to be able to legitimately 'combine checking and trusting' (Power, op. cit., p. 2) and more had to be known about human judgement processes. To help narrow the gap between the high and low grounds of curriculum assessment and clinical practice, this study sought to discover the pattern of judgement followed by supervisors and explore how their expert judgement could be restored through the framework of CiPs.

The literature helped to define the new responsibilities set out for the role of supervisor (page 87). From the accounts of participants taking part in this study, the theory of Judgement in Action emerged and provided a view of how supervisors tended to judge trainee performance and progression. Judgement in Action did this through an innovative adaptation of Schön's reflection in action (1990). According to the theory of Judgement in Action, supervisors made situational judgements in the moment, rather than through competence checklists. Judgement in Action had a unique quality, showing that the capabilities that supervisors looked for in trainees fell into four capability categories (competence, decision-making, administration and relationships). These four capability types were necessary regardless of the training situation being assessed.

Having addressed the first aim of the study, determining how supervisors judged trainee performance and progression through Judgement in Action, the second aim was to determine whether and how the new CiPs framework might have altered and/or resembled the supervisor approach. Judgement in Action and CiPs had common elements including stressing the importance of the knowledge supervisors gained about trainees through the training relationship, the employment of supervisor professional opinions and the use of scaffolding where the lessening of supervision resulted from increased trainee proficiency. However, the study also showed that the increased formality of the CiPs assessment meant that CiPs could seem less attractive to supervisors than the current competence-based approach. In addition, CiPs had a duality of purpose; the first purpose was to help trainees learn and develop (formative) and the second was to provide evidence for judgements on their progression (summative). A tension could result from the challenge posed by combining these two purposes (GMC, 2011). Clinical Supervisors and trainees would need to be able to distinguish between the use of CiPs to provide feedback on progress and the marking of CiPs as a final assessment on the basis that the former contributed to the latter.

Following this study, three options, which might improve CiPs through the application of the theory of Judgement in Action, presented themselves. It also provided a framework upon which supervisors and trainees might reflect on judgement practices. Additional studies would help to confirm whether or not CiPs could successfully adopt the elements of Judgement in Action to meet the twin aims of drawing out a more rounded picture of trainees using supervisor judgement and bridging the gap between curriculum design and surgical practice.

If so, the application of research-based theory may be aligned with the problems of greatest human concern.

Appendix 1: Glossary

Term	Abbreviation	Definition
Annual Review of Competence Progression	ARCP	An annual review panel that makes a summative assessment of trainee progress that allows transition to the next level of training.
Capability		Ability or fitness to perform a function safely.
Capabilities in Practice	CiPs	The surgical curriculum's version of EPAs. CiPs are also the high-level curriculum learning outcomes mandated by the GMC. Each of the nine CiPs covers a key activity of a surgeon, setting out relevant knowledge, clinical skills and professional behaviours relevant to that activity. Trainees must be able to perform all nine CiPs independently (without need of supervision) to be awarded certification, apply for consultant posts and be listed on the GMC's specialist register.
Clinical Supervisor / Supervisor		A consultant surgeon with supervisory responsibility for trainees - either for day-to-day training or overall educational progression.
Competence		The ability to do something successfully or efficiently.
Competence-based assessment	CBA	Assessment systems within curricula that aimed to clearly define the competencies that needed to be achieved and assessed by breaking down tasks into constituent elements to evidence the presence or non-presence of ability.
Curriculum Developers		The ISCP Surgical Director, Head of ISCP, ISCP team, reporting to the ISCP Management Committee and the Joint Committee of Surgical Training (JCST).
Entrust		See trust.

Entrustable Professional Activities	EPA	Units of professional practice, defined as tasks or responsibilities to be entrusted to the unsupervised execution by a trainee once he or she has attained sufficient specific competence (Ten Cate, 2013).
General Medical Council	GMC	The statutory body responsible for setting the standards for postgraduate medical education and for approving curricula against these standards.
Generic		Applicable to all trainees regardless of specialty, discipline and level of training, e.g. generic professional capabilities.
Generic Professional Capabilities	GPC	A syllabus framework of professional behaviours created by the GMC and Academy of Medical Royal Colleges that underpin the medical professional practice of all doctors.
Performance		Doing a task competently (the outcome of competence).
Professional skills		See Generic Professional Capabilities.
Progression		Readiness to practice (allied to judgement of supervision levels).
Schools of Surgery		Regional postgraduate medical education bodies responsible for overseeing the delivery of all aspects of training in the ten recognised surgical specialties and core training.
Supervision level		The level of supervision required by a trainee to undertake a CiP, levels range from ‘able to observe only’; ‘able to act with direct supervision’; ‘able to act with indirect supervision’; ‘able to act unsupervised’ and ‘demonstrates performance to a level well beyond that expected of a day-one consultant’.
Surgical specialties		Ten separate surgical specialisms that trainees can elect to follow after a common core phase.
Trainees		Surgeons in training. Mature postgraduate surgeons who are qualified with a medical degree and practising within a surgical training programme. These individuals are not considered to be fully trained until they complete the programme and are certified for consultant practice, a process which takes at least ten years from completing medical school.

Training placement		A work placement in a surgical specialty unit under a supervisor and within a team to gain experiential learning
Training programme		A rotation of placements in different units in a specialty or mix of allied specialties.
Trust		The granting of permission to perform a function associated with patient safety.
Workplace-based assessment	WBA	Assessments involving the supervisor observing the trainee carrying out a particular area of clinical practice in the clinical setting for the purpose of providing structured feedback (see also appendix 9).

Appendix 2: Study participants

Pseudonym	Specialty	Region	Experience (years)	Gender	Background
1	General Surgery (Cancer)	West Midlands	3	Male	Non-UK
2	Paediatric Surgery	Northern	5	Female	UK
3	Plastic Surgery	London	4.5	Male	UK
4	General Surgery (HPB/UGI)	Scotland	3	Male	UK
5	General Surgery (Breast)	Yorkshire & Humber	15	Female	UK
6	Neurosurgery	Yorkshire & Humber	8	Male	Non-UK

Appendix 3: Interview Schedule

---

1. Pre-exercise interview - judgement

Exploratory and open. Prompting stories. The following were prompts for the researcher.

Warm up

How do you find being a Clinical Supervisor?

Barriers / Environment / Influences

What do you find makes it easier to be a CS?
What do you find makes it more difficult to be a CS?

Judgement

Can you think of 2 of your current trainees who are at different levels and describe a few things for me:

- What levels are they?
- How would you summarise their performance?
- Tell me about the gap between where each of them is now and where they should be as a day one consultant.

From when you first meet your trainees to the time they leave you, how do you develop your knowledge of what they're capable of? Stories

For you, what sort of information helps you know you've judged your trainees' right?

If you could go back to the time when you first became a trainer, what advice would you give to yourself about how to judge trainees – what to look out for? Stories

(*any pitfalls to avoid?)*

I suppose there's always a first time for a trainee to do something on their own.

Thinking about your current trainees, can you give me an example of a time when they were allowed to do something on their own that they'd never carried out before on their own?

What was it?

How did you know they could be trusted to do that?

How do you know that they did it ok?

What standard do you judge them against?

Would you be happy for them to operate on you? Why / Why not?

What are the factors that make you feel you need to monitor a trainee more or less?

Are there any things that you feel interfere with you trusting trainees to do more than they do now?

(observed them, spoken, seen work, checked, previous, time with them / relationship / patient)

Every trainee is different - how do you manage those who are slower / high-flyers?

In what circumstances do you confer with anyone else?

Further prompts if necessary:

<u>*Clinical assessment*</u>

- *What are they like when they're talking to patients 1:1?*
- *Would you be happy if the patient was your Mum?*
- *Relaxed?*
- *Command of the background?*
- *Can they apply their knowledge?*
- *How focused are their approaches to patient management?*
- *Choice of investigation - can they narrow things down?*
- *Diagnosis*
- *How well do they multi-task?*

<u>*Communication and team*</u>

- *What is their communication like with their peers? Mature?*
- *What are their written letters like?*
- *How comfortable are they with members of staff?*
- *How comfortable are they with being the centre of attention e.g. leading a ward round?*
- *How are they at the MDM meeting?*
- *How do they handle conflict?*

<u>*Situational awareness*</u>

- *What's their time management like?*
- *Do they recognise the impact of time on the delivery of care?*

<u>*Decision-making*</u>

- *How decisive are they? Prepared to make decisions?*
- *How complete a picture do they need in order to come to the correct decision?*
- *How well can they deal with uncertainty?*
- *How do they respond when under pressure or stress?*

<u>*Leadership*</u>

- *How good are they at orchestrating a whole team e.g. in theatre? Multi-faceted team role*

Theatre

2. Post-exercise interview – revised judgement

Exploratory and open. The following were prompts for the researcher.

Decision making Talk me through your thinking on each CiP (Taking each trainee / each CiP in turn) How would you judge them in relation to the CiP? What supervision level would you give / why? How do you think this process compares with how you judge your trainees now? What difference did the CiPs make to you if any? Any surprises?
Use and content Is there anything about these CiPs that you would change? Was anything missing? If you had 9 of these CiPs to do on each of your trainees, how would you go about it? Would you prefer doing it on your own or in a group *how would that work?* What will work and not work? (*Why*? *What might help resolve these*) What would need to be done to implement this successfully?

Tasks	Knowledge and skills
1. Assesses and prioritises GP and inter-departmental referrals 2. Arranges pre-clinic investigations 3. Gives advice when requested and appropriate 4. Deals correctly with inappropriate referrals 5. Assesses new and review patients, determines the appropriate plan of action, explains to patient and executes. Includes: a) discharge with advice b) further review with or without investigation c) list for surgery d) refer onwards for further opinion e) admit for urgent investigation and / or treatment 6. Communicates consultation outcomes and subsequent investigation results by appropriate means 7. Completes all necessary documentation 8. Manages time with the clinic setting 9. Observes appropriate infection control protocols 10. Delegates and trains on appropriate cases	1. Applies syllabus defined knowledge and skills in straightforward and unusual cases across the breadth of the specialty 2. Uses a structured history and a focussed clinical examination to perform a full clinical assessment 3. Carries out syllabus defined practical investigations or procedures within the out-patient setting 4. Adapts approach to accommodate all channels of communication (e.g. interpreter, sign language) and to appropriately involve relatives and friends. 5. Synthesises clinical findings into an overall impression and a differential diagnosis 6. Identifies patients with unusual, serious or urgent conditions 7. Takes co-morbidities into account 8. Requests appropriate investigations and does not investigate when not necessary 9. Interprets results of investigations in context 10. Exercises good judgement in deciding on management plans and executes these within appropriate timescales 11. Effectively manages potentially challenging situations in patients with complaints 12. Emphasises health promotion

Appendix 5: Capability in Practice 3 - Managing Ward Rounds and On-going Care of In-Patients

Tasks	Knowledge and skills
1. Identifies at the start of a ward round if there are acutely unwell patients who require immediate attention 2. Ensures that all necessary members of the multi-disciplinary team are present, know what is expected of them and what each others' roles and contributions will be 3. Ensures that all documentation (including results of investigations) will be available when required 4. Manages time well, ensuring all necessary assessments are made and discussions held, while continuing to make progress with the overall work load of the ward round 5. Carries out the following activities, as appropriate, for each patient and tailoring the approach to specific patients as required: a) Greets the patient and explains the discussions and activities that will take place on the ward round b) Ensures that all members of the multi-disciplinary team are aware of the relevant history and clinical course to date c) Speaks to and examines the patient according to the clinical situation d) Observes infection control protocols and ensures other team members are compliant also e) Ensures input is obtained from all members of the multi-disciplinary team as appropriate and takes this into account when formulating a management plan f) Ensures the results of all relevant investigations are seen and taken into account when formulating a management plan g) Identifies when other factors (e.g. co-morbidity, family or social circumstances) are of relevance and pays due regard to these h) Recognises when advice from other specialties is required, and ensures this advice sought and acted on within the overall context for an individual patient i) Uses good judgement to decide on appropriate management plans for each patient and communicates these clearly to the patient and to the multi-disciplinary team j) Ensures appropriate documentation of the findings and management plan in the case notes 6. Recognises when further therapeutic manoeuvres are not in the patient's best interests, initiates palliative care, refers for specialist advice as required and discusses plans with the patient and their family 7. Summarises important points at the end of the ward round and ensures all members of the multi-disciplinary team understand the management plans and their roles within them 8. Works within a 'consultant of the week' model of care or equivalent.	1. Applies syllabus defined knowledge and skills in straightforward and unusual cases across the breadth of the specialty 2. Makes full assessment of patients clinically across the breadth of the specialty, focussing the history and clinical examination as required 3. Recognises when the clinical course is progressing as expected and when medical or surgical complications are developing 4. Identifies and initially manages co-morbidity and medical complications, referring on to other specialties as appropriate 5. Contributes effectively to level 2 and level 3 care 6. Demonstrates the ability to contribute effectively to cross specialty working 7. Synthesises findings into an appropriate overall impression of the patient's progress, making diagnoses as required and identifying when the patient has an unusual, serious or urgent condition 8. Uses good judgement to decide on a detailed management plan and ensures this is executed within the appropriate timescale 9. Gives appropriate advice for discharge documentation and follow-up 10. Recognises when operative intervention or re-intervention is required and ensures this is carried out 11. Manages potentially difficult or challenging interpersonal situations, including complaints 12. Uses the consultation to emphasise health promotion to patients

Tasks	Knowledge and skills
1. Selects patients appropriately for surgery, taking the surgical condition, co-morbidities, medication and investigations into account 2. Discusses reasonable treatment options and shares decision making with patients 3. Takes informed consent or applies national legislation for patients who are not competent to give consent 4. Arranges anaesthetic assessments as required 5. Undertakes the appropriate process to list the patient for surgery 6. Prepares the operating list, accounting for case mix, skill mix, operating time, clinical priorities and patient co-morbidity 7. Leads the brief and ensures all relevant points are covered for all patients on the operating list 8. Ensures the WHO pre-operative checklist is completed for each patient 9. Carries out the operative procedures as required in the specialty syllabus 10. Maintains overall situational awareness throughout the operating list and reacts appropriately 11. Ensures that all members of the team respect the sterile operating field and the risk of contamination is minimised. 12. Interacts appropriately with all members of the multi-disciplinary team 13. Ensures the WHO end of procedure checklist is carried out for each patient 14. Writes a full operation note for each patient, ensuring inclusion of all post-operative instructions 15. Leads the de-brief, ensuring all relevant points are discussed 16. Completes any appropriate logbook, audit or database entries 17. Reviews all patients post-operatively	1. Applies the legal requirements to the taking of consent and has a good understanding of when a patient is competent to give consent 2. Leads the preoperative briefing 3. Applies syllabus defined knowledge and skills in straightforward and unusual cases across the breadth of the specialty 4. Synthesises the patient's surgical condition, the technical details of the operation, co-morbidities and medication into an appropriate operative plan for each patient 5. Carries out the operative procedures as described in the syllabus 6. Uses good judgement to adapt operative strategy to take account of pathological findings and any changes in clinical condition 7. Undertakes the operation in a technically safe manner, using time efficiently 8. Understands when prophylactic antibiotics should be prescribed, and which antibiotics have a suitable bactericidal action allowing them to be used for prophylaxis 9. Manages complications safely, requesting help from colleagues where required. 10. Demonstrates good application of knowledge and non-technical skills in the operating theatre, including situation awareness, decision making, communication, leadership and teamwork

Generic Professional Capabilities	
Domain 1: Professional values and behaviours • Maintains trust, shows respect, courtesy, honesty, compassion, empathy • Treats patients as individuals, respecting dignity and confidentiality • Takes prompt action for patient safety • Manages time and resources and health and safety compliance • Self-monitors and seeks appropriate advice • Has emotional resilience • Demonstrates situational awareness including human factors • Reflects on personal behaviour and awareness of own behaviour • Demonstrates awareness of own limitations and seeking help where required • Interacts with colleagues • Works towards a safe and supportive working environment • Has a commitment to learn and continuing professional development • Ability to accept constructive feedback • A professional role model **Domain 2: Professional skills** **Practical skills** • Articulate, literature, numerate, good English, IT skills • Gives clear and accurate information and instructions • Makes clear written records • Has knowledge of legal aspects and information governance **Communication and interpersonal skills** • Shares decision making, respecting the patient's concerns and expectations • Demonstrates cultural and social awareness • Communicates effectively and sensitively • Works with team to re-assess priorities and manage complex situations • Demonstrate safe and effective handover • Ability to apologise	**Dealing with complexity and uncertainty** • Judgement and ability to manage uncertainty and make reasonable adjustments • Resilience, diligence and thoroughness • Recognise limits of own competence and seek help where necessary **Clinical skills** • History taking, examination skills and clinical reasoning • Select and interpret investigations, safe operation of medical devices • Takes consent, safe prescribing **Domain 3: Professional knowledge** • Keeps up to date • Understands risk • Takes part in audit, appraisal, planning. **Domain 4: Capabilities in health promotion** • Understands factors relating to health inequalities • Promotes health and well-being • Understands risks and interventions **Domain 5: Capabilities in leadership and team working** • Demonstrates leadership and ability to adapt to improve outcomes. • Actively participates to the success of the team • Thinks critically about decision making and explains decisions transparently • Team-working, appreciating the roles of all members of the team **Domain 6: Capabilities in patient safety and quality improvement** • Raises safety concerns appropriately • Understands multi-disciplinary roles • Promotes patient involvement • Gives effective briefing and debriefing and makes accurate records **Domain 7: Capabilities in safeguarding vulnerable groups** • Knows legislation, identifies risks, escalates concerns **Domain 8: Capabilities in education and training** • Takes part in effective education training **Domain 9: Capabilities in research and scholarship** • Keep up to date with and apply research best practice • Critically appraise research literature

Appendix 8: Coding table

Categories	Codes and sub-codes		P1	P2	P3	P4	P5	P6
Supervisor factors	**Supervisor control**							
	1. Being present / ready to take over		4			3		5
	2. Task / unit ownership		7	1		6	2	4
	3. Patient ownership					3	1	2
	4. Supervisor role		2	3	5	2		
	5. Information from source • Trainee • Assessment • Outcomes • MPT • Other trainers		See sources of evidence below					
	Supervisor trust							
	• Remove protection • Let trainees do it their own way		1	3		2		
	Personal factors							
	1. Culture		1					
	2. Ability		1		1			
	3. Biases • Difficulty being objective • Difficulty with different cultures • Favour trainees they know well				2			1
	4. Personal rules		1			1		1
	5. Trainer feelings of competence			1	2	1	1	
	6. Pragmatic approach		2					
	7. Specialty differences • Less meticulous specialty • Specialty of rarities • Lack of understanding for specialty • All surgeons need fast thinking		1	1				3
	8. Relationship with other supervisors (see trainee qualities below)		2	8		1		

	9. Past training experience • Cultural differences • Overcoming skill deficiencies • Lack of structure / numbers • Being allowed to do more / no supervision • Competition as a female / part-time • Commitment to career vs job and home • More time on the job • Not being allowed to call the consultant		2	1			1	5
Judgement in action	**Judgement**							
	1. Nature of supervision							
	• Direct observation / supervisor present		4	1	1		1	1
	• Work nearby trainees		3	1		1		
	• Directly assess		1	1			1	1
	• Take over if struggling		1					1
	• Let trainee do parts		2	1	1	2		
	• Incremental - observe, assist, assisted, do straightforward, do whole, teach		2	1		2	4	1
	• Discuss with trainee		1		1	1		
	• Use safest materials							1
	• Give trainee appropriate cases			1		1		
	• Reliance on nurse supervision		1					
	2. How you know to trust							
	• Hand skills		3			4		1
	• Ability to synthesise patient case		1					2
	• Ability to ask / call supervisor		2		1	1	2	2
	• Trainee confidence / focus		1	1	1			1
	• Trainee can perform / direct proceedings		1	1	2	2	2	1
	• Trainee progressing at right rate		2		3			
	• Trainee able to think, decide, plan			3	3	2		1
	• Other trainers' positive experiences			2	1	2		
	• Trainee's previous jobs			2	1			
	• Workplace assessment			1				

	• Direct observation / checking		2	4		4		1
	• Insight about ability				1	1	3	1
	• Trainee outcomes				2			
	• Experience / rarities / complications			2		1	2	
	• Ability to progress as well as keep safe					1		
	• Trainee proactive					1		1
	• Stage of training					1		
	• Trainee talks through how they would do it					1		1
	• Question their applied / knowledge						1	3
	• Get to know trainee		1	1		2	1	1
	• To keep trainee from disengaging				1	1		1
	3. Reasons for first time decision							
	• Know trainee ability		1	1	3	2	2	3
	• Gradual process		1	2		1		1
	• Trainee experience		1	1		1		1
	• Trainee knowledge							2
	• To motivate		2	1		1		1
	• Right for stage of training / peer level		1	3	1			
	• Negotiated with trainee						1	
	• If trainee involved with patient			1				2
	• Skills proven in different procedure			1				
	• Time pressures				1			
	• Trainee has insight into own ability				1			
	• Carefully planned, checked with supervisor					1		
	• Supervisor nearby					1		
	• Trainee a high flier					1		
	4. When you know to trust							
	• At incision, positioning, assisting, closing		1			1		1
	• When know location of previous training						1	
	• When talk / get to know them			1			2	2

	• When confident						1	1
	• When committed / proactive / honest						1	2
	• Not before 8 weeks in				1			
	• Longer because of shift patterns		2					1
	5. What might not be observed							
	• Rare procedures			3				
	• Behaviour with other people			1	1			
	• Ward rounds					1		
	• Some operations					1		
	• Clinics						1	1
	• Acute work						1	
	• Trainee being overstretched						1	
	• Seeing trainees on a regular basis						1	
	6. Saying no to trainees							
	• When trainees have not progressed patient			1				1
	• When trainee is new			1				
	• Patient very sick / difficult op / parent issues			1				1
	• Trainee issues of knowledge, competence						1	
	7. Gut feeling					2		3
	Standard							
	1. Capable for stage of training				1	5	3	
	2. Patient safety / care		1	1		2		5
	3. Protocols						2	
	4. Able to do tasks / procedures					1		2
	5. Textbook		1					
	6. Self / colleagues		1			1		2
	7. Curriculum			1				
	Sources of evidence for making judgements	**Trainee**						
		1. Trainee reports about patients			1			
		2. Trainee view (initial trust)			1			

		Assessment						
		1. Direct observation (grounded trust)	3		3			2
		2. Assessments by others / grades (presumptive trust)	3				2	
		3. Multi-source feedback assessment		3			1	
		4. Assessments by participant (grounded trust)	1					
		Outcomes						
		1. Unsatisfactory results			2			
		2. Quality of clinic letters	1				1	
		3. Electronic system data	1				1	
		4. Focus of investigations ordered	1		1			
		Judgement of MPT						
		1. Nurse report	4	3	3		3	
		2. Who juniors call first	1					
		3. Patient reports			1		2	
		4. Team			1		3	2
		5. GPs					1	
		Judgement of other supervisors						
		1. Information from other supervisors	4	1	4	4	3	7
	Trainee qualities – positive	**Ability**						
		1. Technical ability / safe	5	3	1	3	2	1
		2. Ability to act independently	3	2	1	2	3	3
		3. Investigations	1			1	1	
		4. Lack of mistakes / complaints	3			1		1
		5. Focused / single-minded	2			2		2
		6. Presenting / writing	1			1	1	1
		7. Prioritising	1	1			1	

		8. Calm under pressure		1				
		9. Confidence	2	2	1	1	1	3
		10. Insight / reflection / ask for help	1	3	3	2	3	4
		11. Handling complex problems	2	3		2	3	4
		12. Taking / maximising time appropriately	3	1				2
		13. Curious / goes out of way to learn	2	1	1	1	2	
		14. Taking responsibility	2	1		1	1	3
		Lessening supervisor workload						
		1. Technical speed	2		1		3	
		2. Proactive	1	1	1	1		2
		3. Organisational	6		1	1		3
		4. Communication – asking / informing	4		1	1	2	3
		5. Synthesising	1	1				
		6. Working hard	2	2	1	2	1	6
		7. Engaged / Committed / Reliable	1	1	2	1	2	6
		8. Knowledge	1	1		2	1	2
		9. Experience		3			7	1
		10. Broadness						1
		11. Probity / Honesty				1		3
		Quality of trainee decisions						
		1. Ability to make decisions (simple to complex)	1	3	1	1	2	3
		2. Decision speed / thinking speed	1			2	1	4
		3. Based on insufficient information		2		2	1	3
		4. Awareness of impact of decisions	1	2			1	2

		5. Appropriately call the consultant	3	3		1	2	5
		6. Sound management plans	2	1	1	1	2	3
		7. Getting it		3			1	3
		Trainee relationships						
		1. With juniors	3	2		2	1	2
		2. With MPT		7	5		9	2
		3. With patients		1		1	1	3
		4. Social / cultural skills – fitting in	1	4	2		2	1
		5. Appropriately competitive	1	1		1		1
		Trainee-supervisor relationship						
		1. Time with trainee				1		1
		2. Commitment		1				2
		3. Knowledge of trainee	1		1		1	4
		4. Apprenticeship			1			1
	Trainee qualities – negative							
	1. Slow / lack of technical skill		3	3			2	
	2. Lack of knowledge / evidence base		1			2	1	3
	3. Over-confidence / arrogance		3	1		2	2	2
	4. Lack of confidence		2	2		1	2	3
	5. Lack of insight		2	1		1	1	3
	6. Leaping ahead					2	1	
	7. Performativity of trainee approach				1			1
	8. Lack of commitment		3	2		1	3	3
	9. Lack of social skills		1	2			3	1
	10. Leaving prematurely		2	1			1	
	11. Different training regimes		1			1		
	12. Not suited to surgery			2		1		3
	13. Unable to make decisions		1	1	1	2	2	4
	14. Completing medical notes					1		

	15. Lack of time management				1	1	1	
	16. Mistakes				1			
	17. Conduct				1			
	18. Cannot take criticism					1		
Nature of clinical decisions	**Expert**							
	1. Speed	2	3	1	1	7	2	
	2. Knowledge / experience – evidence base rarities / variety / complexity	1	3	1	3			
	3. Decisions with little info / uncertainty	1	2	1	1	1	2	
	4. Know when to consult / re-think	3	3		1	2	5	
	5. Sensitivity to culture / team / patients			1				
	6. Open mind / options				1			
	7. Different ways to same end	2		1				
	8. Prioritising patient need	2	1		1			
	9. Relationship with other supervisors	7	8	1	2	2	2	
	10. Gravitas of title					1		
	Environment							
	1. Clinic	4	1			1	3	
	2. Theatre	4	1	1	1	3	2	
	3. Ward	1						
Learning	1. Ipsative, trainee differences	6	9	4	8	6	5	
	2. Non-ipsative, standard pathways		4		2	2	2	
	3. Use of assessment	7	2	1	1	2		
	4. Learning from all opportunities	2	3			5	2	
	5. Reward / motivation (i.e. allowing trainees to operate)		1	1	1		2	
	6. Need for trainee vocation				2		1	
	7. Zone of proximal development e.g. telling trainee in advance to allow mental rehearsal		3		1	1	2	
	8. Modular (learn bit by bit)	1	6		3	4		

Constraint	**Time**		11	5		1	3	4
	Service							
	1. Overbooking / Targets / Money		9	4	1	1	1	3
	2. Conflict with training		5	2			3	2
	3. Fracturing work			4		1	2	1
	4. Gaps in training / lack of opportunities		4	3	1			1
	Performativity							
	1. Assessment and feedback		3	7	6	1	2	
	2. The system			2	1	1		
	3. Need for evidence			9				1
	4. Trainee / trainer lack of commitment		1	1	1			4
	Trainee cultural differences		1	8		2	1	

Appendix 9: Workplace-based assessment methods (WBA)

Method	Abbreviation	Description
Assessment of Audit	AoA	Assesses competence in completing an audit or quality improvement project based on the review of project documentation or a presentation at an audit meeting.
Case Based Discussion	CBD	Assesses clinical judgement, decision-making and the application of medical knowledge in relation to the management of a challenging patient case and includes the ability to reflect on practice. It includes a structured in-depth discussion between trainee and consultant supervisor, using the patient records as a basis for the dialogue.
Direct Observation of Procedural Skills in Surgery	DOPS	Assesses competence in technical, operative and professional skills in a range of basic diagnostic and interventional procedures during routine surgical practice.
Multi-Source Feedback	MSF	Assesses the ability to work professionally as part of a multi-disciplinary team. It comprises a trainee self-assessment and assessments from a range of the trainee's co-workers, including the educational supervisor. It culminates in a personalised and anonymised chart showing a comparison of trainee/team views and a meeting between the trainee and educational supervisor to discuss the feedback.
Observation of Teaching	OoT	Assesses instances of formal teaching delivered by the trainee.
Procedure Based Assessment	PBA	Assesses technical, operative and professional skills in a range of specialty / advanced procedures during routine surgical practice.

References

Academy of Medical Royal Colleges & General Medical Council. (2017). Generic professional capabilities: guidance on implementation for colleges and faculties.

Alvesson, M., Hardy, C. and Harley, B. (2008). 'Reflecting on reflexivity: reflexive practices in organization and management theory'. *Journal of Management Studies*, 45, 480–501.

Anderson, N. H., & Lopes, L. L. (1974). Some psycholinguistic aspects of person perception. *Memory & cognition*, *2*(1), 67-74.

Ball, S. J. (2008). Performativity, privatisation, professionals and the state.

Barber, B. (1963). Some problems in the sociology of the professions. *Daedalus*, 669-688.

Baxter, P., & Jack, S. (2008). Qualitative case study methodology: Study design and implementation for novice researchers. *The qualitative report*, *13*(4), 544-559.

Bazeley, P. & Jackson, K. (2013) *Qualitative Data Analysis with NVivo*. SAGE.

Beck, U. (1992). *Risk society: Towards a new modernity* (Vol. 17). Sage.

Blank, W. E. (1982). *Handbook for developing competency-based training programs*. Englewood Cliffs, NJ: Prentice-Hall.

Bleakley, A. (1999). From reflective practice to holistic reflexivity. *Studies in higher education*, *24*(3), 315-330.

Bloom, B. S. (1971). *Taxonomy of Educational Objectives: The Classification of Educational Goals: By a Committee of College and University Examiners*. David McKay.

Bloxham, S. (2009). Marking and moderation in the UK: false assumptions and wasted resources. *Assessment & Evaluation in Higher Education*, *34*(2), 209-220.

Bolger, F., & Wright, G. (1994). Assessing the quality of expert judgment: Issues and analysis. *Decision support systems*, *11*(1), 1-24.

Braun, V. & Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology*. [Online] 3 (2), 77–101. Available from: doi:10.1191/1478088706qp063oa.

Brown, M., Patrick, F., Tate, D., & Wright, S. (1994). A collection of readings related to competency-based training. *Victoria, Australia: Victorian Education Foundation, Deakin University*.

Burr, V. (2003) *Social constructionism*. [Online]. Psychology Press. Available from: http://books.google.co.uk/books?hl=en&lr=&id=AR_lQURihrgC&oi=fnd&pg=PP11&dq=vivien+burr&ots=UDc41Bzs6h&sig=TLOecLtcB91Eihr55xwLs86lNUM [Accessed: 11 April 2015].

Bussey, M. (2016). Qualitative analysis of written formative feedback in core surgical training. *British Journal of Surgery*, *103*, 46-47.

Bussey, M. (2017), *The Feedback FREND: The development and pre-pilot evaluation of a trainee-led feedback tool for surgical training.* Institutional-Focused Study, UCL.

Bruner, J. (1984). Vygotsky's zone of proximal development: The hidden agenda. *New Directions for Child and Adolescent Development*, *1984*(23), 93-97.

Camerer, C. F., & Johnson, E. J. (1997). The process-performance paradox in expert judgment: How can experts know so much and predict so badly. *Research on judgment and decision making: Currents, connections, and controversies*, *342*.

Choo, K. J., Arora, V. M., Barach, P., Johnson, J. K., & Farnan, J. M. (2014). How do supervising physicians decide to entrust residents with unsupervised tasks? A qualitative analysis. *Journal of hospital medicine*, *9*(3), 169-175.

Coles, C. (2002). Developing professional judgment. *Journal of Continuing Education in the Health Professions*, *22*(1), 3-10.

Collins, M. (1983). A critical analysis of competency-based systems in adult education. *Adult education*, *33*(3), 174-183.

Costley, C., Elliott, G., & Gibbs, P. (2010). Key concepts for the insider-researcher. *Costley, C., Elliott, G. & Gibbs, P. Doing work based research: Approaches to enquiry for insiderresearchers*, 1-7.

Crook, D. (2008). Some historical perspectives on professionalism.

Coulson, A. (2009). Targets and terror: Government by performance indicators. *Local Government Studies*, *35*(2), 271-281.

Croskerry, P., Petrie, D. A., Reilly, J. B., & Tait, G. (2014). Deciding about fast and slow decisions. *Academic Medicine*, *89*(2), 197-200

Crossman, A. (2018). The major theoretical perspectives of sociology. *ThoughtCo.*<https://www.thoughtco.com/theoretical-perspectives-3026716

Cunningham, B. (2008) *Exploring professionalism*. [Online] Available from: http://philpapers.org/rec/CUNEP [Accessed: 20 February 2018].

Damodaran, A., Shulruf, B., & Jones, P. (2016). 'Trust 'versus 'competency' in the workplace. *Medical Education*.

Denscombe, M. (2010). *Ground rules for social research: Guidelines for good practice*. Open University Press

Department of Health (2014) Modernising Medical Careers. *Wikipedia, the free encyclopedia*. [Online]. Available from: http://en.wikipedia.org/w/index.php?title=Modernising_Medical_Careers&oldid=631064884 [Accessed: 14 December 2017].

Department of Health, E. (2014). The Belmont Report. Ethical principles and guidelines for the protection of human subjects of research. *The Journal of the American College of Dentists*, *81*(3), 4.

Dewey, J. (1916). Democracy and Education, New York. *Macmillan). If you see john Dewey, tell him we did it. Educational Theory*, *37*(2), 145-152.

Dewey, J. (2004). *Democracy and education*. Courier Corporation.

Dijksterhuis, M. G., Voorhuis, M., Teunissen, P. W., Schuwirth, L. W., Ten Cate, O. T., Braat, D. D., & Scheele, F. (2009). Assessment of competence and progressive independence in postgraduate clinical training. *Medical education*, *43*(12), 1156-1165.

Dino, G. A., Shanteau, J., Binkley, M., & Spenser, A. (1984). *The detrimental effects of environmental stress on creativity*. Tech. Rep. 84–2). Manhattan: Kansas State Univer-sity, Department of Psychology.

Donaldson, S.L. (Chief M.O. (2002) *Unfinished business: proposals for reform of the senior house officer grade - a paper for consultation*. [Online]. 21 August 2002. Available from: http://webarchive.nationalarchives.gov.uk/+/www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4007842 [Accessed: 28 December 2014].

Dreyfus, H. L., & Dreyfus, S. E. (2005). Peripheral vision: Expertise in real world contexts. *Organization studies*, *26*(5), 779-792.

Edwards, A. D., & Furlong, V. J. (1985). Reflections on the language of teaching. *Field Methods in the Study of Education, Falmer Press, London*, 21-36.

Eliot, T. S. (1909). Choruses from "The Rock" 1934. *Collected poems*, *1962*, 145-171.

Eraut, M. (1994). *Developing professional knowledge and competence*. Psychology Press.

Eraut, M. (1995). Schön Shock: a case for refraining reflection-in-action?. *Teachers and teaching*, *1*(1), 9-22.

Eraut, M. (2000). Non-formal learning and tacit knowledge in professional work. *British journal of educational psychology*, *70*(1), 113-136.

Ericsson, K. A., & Smith, J. (Eds.). (1991). *Toward a general theory of expertise: Prospects and limits*. Cambridge University Press.

Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic medicine*, *79*(10), S70-S81.

Ericsson, K. A., Prietula, M. J., & Cokely, E. T. (2007). The making of an expert. *Harvard business review*, *85*(7/8), 114.

Ethical Guidelines for Educational Research 2011 | *BERA*. [Online]. Available from: https://www.bera.ac.uk/researchers-resources/publications/ethical-guidelines-for-educational-research-2011 [Accessed: 15 May 2015b].

Feigenbaum, E., & P. McCorduck (1983). *The Fifth Generation: Artificial Intelligence and Japan's computer challenge to the world.* Reading, MA: Addison-Wesley.

Few, S. (2015). *Signal: Understanding what Matters in a World of Noise* (Vol. 1). Burlingame, CA: Analytics Press.

Fish, D., & Coles, C. (1998). Developing professional judgement in health care.

Franzese, C.B. & Stringer, S.P. (2007) The Evolution of Surgical Training: Perspectives on Educational Models from the Past to the Future. *Otolaryngologic Clinics of North America*. [Online] 40 (6), 1227–1235. Available from: doi:10.1016/j.otc.2007.07.004 [Accessed: 21 December 2014].

Fraser, S. W., & Greenhalgh, T. (2001). Complexity science: Coping with complexity: Educating for capability. *BMJ: British Medical Journal*, *323*(7316), 799.

Freidson, E. (1994) *Professionalism Reborn [electronic resource]: Theory, Prophecy and Policy*. 1st edition. Hoboken, Wiley.

Ganzach, Y. (1997). Theory and configurality in clinical judgments of expert and novice psychologists. *Journal of Applied Psychology*, *82*(6), 954.

Gingerich, A. (2015). What if the 'trust' in entrustable were a social judgement?. *Medical education*, *49*(8), 750-752.

General Medical Council (2011). *Learning and assessment in the clinical environment: the way forward*.

General Medical Council. (2017). Generic Professional Capabilities Framework.

General Medical Council (2016). *Generic professional capabilities - outcome of public consultation.* [Online]. Available at: http://www.gmc-

uk.org/06___Generic_professional_capabilities___published_version.pdf_64818615.pdf

General Medical Council. (2017). Excellence by design: standards for postgraduate curricula.

General Medical Council. (2010). Workplace Based Assessment: A guide to implementation.

Govaerts, M. J. (2017). Trust, entrustment decisions and a few things we shouldn't forget. *Perspectives on medical education*, *6*(2), 68.

Gov.uk. (2010). *Equality act 2010:guidance - GOV.UK*. [online] Available at: https://www.gov.uk/guidance/equality-act-2010-guidance

Gov.uk. (2018). *Data protection - GOV.UK*. [online] Available at: https://www.gov.uk/data-protection/the-data-protection-act [Accessed 10 Dec. 2018].

Hackman, J. R., & Oldham, G. R. (1976). Motivation through the design of work: Test of a theory. *Organizational behavior and human performance*, *16*(2), 250-279.

Harlen, W., & James, M. (1997). Assessment and learning: differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice*, *4*(3), 365-379.

Harris, I. B. (1989). A Critique of Schön's Views on Teacher Education: Contributions and Issues. *Journal of Curriculum and Supervision*, *5*(1), 13-18.

Hauer, K. E., Oza, S. K., Kogan, J. R., Stankiewicz, C. A., Stenfors-Hayes, T., Cate, O. T., ... & O'sullivan, P. S. (2015). How clinical supervisors develop trust in their trainees: A qualitative study. *Medical education*, *49*(8), 783-795.

Hauer, K. E., Kohlwes, J., Cornett, P., Hollander, H., Ten Cate, O., Ranji, S. R., ... & O'Sullivan, P. (2013). Identifying entrustable professional activities in internal medicine training. *Journal of graduate medical education*, *5*(1), 54-59.

Highlight Reports. (2018, November 13). Retrieved from https://deliveringresults.leeds.ac.uk/pm-support/highlight-reports/A Practical Guide To Delivering Results

Holzhausen, Y., Maaz, A., Cianciolo, A. T., Ten Cate, O., & Peters, H. (2017). Applying occupational and organizational psychology theory to entrustment decision-making about trainees in health care: a conceptual model. *Perspectives on Medical Education*, *6*(2), 119-126.

Hughes, G. (2014). Making the Case for Ipsative Assessment: Assessment for Learning Revisited. In *Ipsative Assessment* (pp. 92-110). Palgrave Macmillan UK.

Ico.org.uk. (2018). *Guide to the General Data Protection Regulation (GDPR)* | *ICO*. [online] Available at: https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/ [Accessed 10 Dec. 2018].

Kirkpatrick, D. L. (1979). Techniques for evaluating training programs. *Training and development journal*.

Janesick, V. J. (2018). The dance of qualitative research design: Metaphor, methodolatry, and meaning. ResearchGate

Johnson, E. J. (1988). Expertise and decision under uncertainty: Performance and process. *The nature of expertise*, 209-228.

Jones Jr, M. D., Rosenberg, A. A., Gilhooly, J. T., & Carraccio, C. L. (2011). Perspective: competencies, outcomes, and controversy—linking professional activities to competencies to improve resident education and practice. *Academic Medicine*, *86*(2), 161-165.

Kelman, H. C. (1958). Compliance, identification, and internalization three processes of attitude change. *Journal of conflict resolution*, *2*(1), 51-60.

Kinsella, E. A. (2007). Technical rationality in Schön's reflective practice: dichotomous or non-dualistic epistemological position. *Nursing Philosophy*, *8*(2), 102-113.

Kolb, D. (1984). Experiential learning as the science of learning and development.

Lewin, K. (1946). Action research and minority problems. *Journal of social issues*, *2*(4), 34-46.

Litzinger, T., Lattuca, L. R., Hadgraft, R., & Newstetter, W. (2011). Engineering education and the development of expertise. *Journal of Engineering Education*, *100*(1), 123-150.

Lunt, I. (2008). Ethical issues in professional life. In Cunningham, B. (2008). Exploring professionalism, 73-98.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of management review*, *20*(3), 709-734.

Miller, GE (1990) The assessment of clinical skills/competence/performance. *Academic Medicine: Journal of the Association of American Medical Colleges*. 65 (9), S63–S67.

Naples, N. A. (1996). A feminist revisiting of the insider/outsider debate: The "outsider phenomenon" in rural Iowa. *Qualitative Sociology*, *19*(1), 83-106.

Neighbour, R. (2004). *The Inner Apprentice: An Awareness-Centred Approach to Vocational Training for General Practice*. 2nd Revised edition. Oxford; San Francisco, Radcliffe Publishing Ltd.

Norcini, J. J. (2003). ABC of learning and teaching in medicine: Work based assessment. *BMJ: British Medical Journal*, *326*(7392), 753.

Norcini, J.J. & Burch, V. (2008). *Workplace Based Assessment as an Education Tool*. Dundee, UK, Association for Medical Education in Europe.

Norris, N. (1991). The trouble with competence. *Cambridge journal of education*, *21*(3), 331-341.

Norris, N. (Ed.). (2014). *Curriculum and the Teacher: 35 years of the Cambridge Journal of Education*. Routledge.

O'Gorman, K., & MacIntosh, R. (2015). Mapping research methods. *Research methods for business and management: A guide to writing your dissertation*, 50-74.

Oxford English Dictionary (2018).

Parker, J. T. (1984). In search of a real analysis: CBAE leaders respond to the Collins critique. *Adult Education Quarterly*, *34*(2), 105-110.

Persaud, R. (2011). *Staying sane*. Random House.

Piaget, J. (1964). Part I: Cognitive development in children: Piaget development and learning. *Journal of research in science teaching*, *2*(3), 176-186.

Power, M. (1999). *The Audit Society: Rituals of Verification*. New Ed edition.

Ratcliff, S. (1984). Rebuttal to: A critical analysis of competency based systems in adult education. *Adult Education Quarterly*, *34*(2), 111-114.

Regehr, G., MacRae, H., Reznick, R.K. & Szalay, D. (1998) Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Academic Medicine: Journal of the Association of American Medical Colleges*. 73 (9), 993–997.

Rethans, J. J., Norcini, J. J., Baron-Maldonado, M., Blackmore, D., Jolly, B. C., LaDuca, T., ... & Southgate, L. H. (2002). The relationship between competence and performance: implications for assessing practice performance. *Medical education*, *36*(10), 901-909.
Reason, J. (1995). Understanding adverse events: human factors. *BMJ Quality & Safety*, *4*(2), 80-89.

Reznick, R.K. (1993). Teaching and testing technical skills. *The American Journal of Surgery*. [Online] 165 (3), 358–361. Available from: doi:10.1016/S0002-9610(05)80843-8 [Accessed: 14 December 2014].

Robson, C. (2011). *Real World Research 3e*. 3rd Edition. Chichester, West Sussex, John Wiley & Sons.

Royal College of Physicians. (2017). Evaluation of the Proof of Concept Study (PoC).

Rush, H. M. (1975). Job design for motivation. Conference Board.

Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, cognition and personality*, *9*(3), 185-211. (Abstract)

Schön, D. A. (1990). *Educating the Reflective Practitioner: Toward a New Design for Teaching and Learning in the Professions*. New Ed edition. San Francisco, John Wiley & Sons.

Shanteau, J. (1988). Psychological characteristics and strategies of expert decision makers. *Acta Psychologica*, *68*(1), 203-215.

Spady, W. G. (1977). Competency based education: A bandwagon in search of a definition. *Educational Researcher*, *6*(1), 9-14.

Sterkenburg, A., Barach, P., Kalkman, C., Gielen, M., & Ten Cate, O. (2010). When do supervising physicians decide to entrust residents with unsupervised tasks?. *Academic Medicine*, *85*(9), 1408-1417.

Swanwick, T. (2005). Informal learning in postgraduate medical education: from cognitivism to 'culturism'. *Medical education*, *39*(8), 859-865.

Talbot, M. (2004). Monkey see, monkey do: a critique of the competency model in graduate medical education. *Medical education*, *38*(6), 587-592.

Temple, J. (2010). *Time for Training: A Review of the impact of the European Working Time Directive on the quality of training*. [Online]. Available at: http://hee.nhs.uk/wp-content/uploads/sites/321/2012/08/Time-for-training-report.pdf.

Ten Cate, O. (2005). Entrustability of professional activities and competency-bases training. *Medical education*, *39*, 1176-1177.

Ten Cate, O. (2006). Medical education: Trust, competence, and the supervisor's role in postgraduate training. *BMJ: British Medical Journal*, *333*(7571), 748.

Ten Cate, O. (2013). Nuts and bolts of entrustable professional activities. *Journal of graduate medical education*, *5*(1), 157-158.

Ten Cate, O., Chen, H. C., Hoff, R. G., Peters, H., Bok, H., & van der Schaaf, M. (2015). Curriculum development for the workplace using entrustable professional activities (EPAs): AMEE guide no. 99. *Medical teacher*, *37*(11), 983-1002.

Ten Cate, O., & Scheele, F. (2007). Viewpoint: competency-based postgraduate training: can we bridge the gap between theory and clinical practice?. *Academic Medicine*, *82*(6), 542-547.

Ten Cate, O., Hart, D., Ankel, F., Busari, J., Englander, R., Glasgow, N., ... & Touchie, C. (2016). Entrustment decision making in clinical training. *Academic Medicine*, *91*(2), 191-198.

Ten Cate, T. J. O., Snell, L., & Carraccio, C. (2010). Medical competence: the interplay between individual ability and the health care environment. *Medical teacher*, *32*(8), 669-675.

Tripp, D. (2011). *Critical incidents in teaching (classic edition): Developing professional judgement*. Routledge.

Tuxworth, E. (1989). Competence based education and training: background and origins. *Competency based education and training*, 9-22.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, *185*(4157), 1124-1131.

Van Der Stuyf, R. R. (2002). Scaffolding as a teaching strategy. *Adolescent learning and development*, *52*(3), 5-18.

Van Der Vleuten, C.P.M. & Schuwirth, L.W.T. (2005) Assessing professional competence: from methods to programmes. *Medical Education*. [Online] 39 (3), 309–317. Available from: doi:10.1111/j.1365-2929.2005.02094.x [Accessed: 14 December 2014].

Vygotsky, L. (1987). Zone of proximal development. *Mind in society: The development of higher psychological processes*, *5291*, 157.

Watling, C., LaDonna, K. A., Lingard, L., Voyer, S., & Hatala, R. (2016). 'Sometimes the work just needs to be done': socio-cultural influences on direct observation in medical training. *Medical education*, *50*(10), 1054-1064.

Wertsch, J. V. (1984). The zone of proximal development: Some conceptual issues. *New Directions for Child and Adolescent Development*, *1984*(23), 7-18.

Wetherell, M., Taylor, S., & Yates, S. J. (2001). *Discourse theory and practice: A reader*. Sage.

Wigton, R. S. (1996). Social judgement theory and medical judgement. *Thinking & Reasoning*, *2*(2-3), 175-190.

Williams, Z. (Presenter), Leonard, P. (Director), & Crabtree, S. (Series Editor) (2017, June 19). *10 things you need to know about the future.* [Television broadcast]. In *Horizon*. BBC2.

Wolf, A. (1995). Competence-based assessment. Chapter 25, in *Competence in the Learning Society*, John Raven and John Stephenson (eds). Peter Lang, New York, pps 453-466. (2001) Reproduced on the HE Academy website by Peter Lang Publishing, Inc.

Wolf, A. (1995). *Competence-based assessment*. McGraw-Hill Education (UK).

Yanow, D., & Tsoukas, H. (2009). What is reflection-in-action? A phenomenological account. *Journal of management studies*, *46*(8), 1339-1364.

Yin, R. K. (2011). *Applications of case study research*. Sage.

Yin, R. K. (2013). *Case study research: Design and methods*. Sage publications.

Secondary literature

Bloom, B. S., & Sosniak, L. A. (1985). Developing talent in young people.

Bruner, J. S. (1996). *The culture of education*. Harvard University Press.

Carr, W. (1995). *For education: Towards critical educational inquiry*. McGraw-Hill Education (UK).

Creswell, J. W. (1998). Qualitative inquiry and research design: Choosing among five designs.

Gonczi, A., Hager, P., & Oliver, L. (1990). *Establishing competency-based standards in the professions*. Canberra: Australian Government Publishing Service.

Grundy, S. (1987). Curriculum: Product or Praxis? Philadelphia: Falmer Press, 1987.

Habermas J. (1972). Knowledge and human interests. London: Heinemann.

Hakim, C. (2012). *Research Design: Succesful Designs for Social Economics Research*. Routledge.

Johnson, P. E. (1983). What kind of expert should a system be? *The Journal of Medicine and Philosophy*, *8*(1), 77-97.

Lave, J. Wenger, E. (1991). Situated Learning: Legitimate Peripheral Participation. Cambridge: Cambridge University Press. Non English

Lipsey, R. (1988). *An art of our own: The spiritual in twentieth-century art*. Shambhala Publications.

Messick, S. (1984). The psychology of educational measurement. *Journal of educational measurement*, *21*(3), 215-237.

Schatzki, T. R. (2005). Peripheral vision: The sites of organizations. *Organization studies*, *26*(3), 465-484.

Stake, R. E. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.

Stenhouse, L. (1975). An introduction to curriculum research and development. (London, Heinemann).

Vickers, G. (1978). Unpublished memorandum, MIT.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge university press.